

The Trans-Ancestral Genomic Architecture of Glycaemic Traits

Ji Chen^{1,2*}, Cassandra N. Spracklen^{3,4*}, Gaëlle Marenne^{1,5*}, Arushi Varshney^{6*}, Laura J Corbin^{7,8*}, Jian'an Luan⁹, Sara Willems⁹, Ying Wu³, Xiaoshuai Zhang^{9,10}, Momoko Horikoshi^{11,12,13}, Thibaud S Boutin¹⁴, Reedik Mägi¹⁵, Johannes Waage^{16,17}, Achilleas Pitsilides¹⁸, Ruifang Li-Gao¹⁹, Kei Hang^{20,21,22}, Jie Yao²³, Mila D Anasanti²⁴, Audrey Y Chu²⁵, Annique Claringbould²⁶, Jani Heikkinen²⁴, Jaeyoung Hong¹⁸, Jouke-Jan Hottenga^{27,28}, Shaofeng Huo²⁹, Marika A. Kaakinen^{30,24}, Tin Louie³¹, Winfried März^{32,33,34}, Hortensia Moreno-Macias³⁵, Anne Ndungu¹², Sarah C. Nelson³¹, Ilja M. Nolte³⁶, Kari North³⁷, Chelsea K. Raulerson³, Debashree Ray³⁸, Rebecca Rohde³⁷, Denis Rybin¹⁸, Claudia Schurmann^{39,40}, Xueling Sim^{41,42,43}, Loz Southam¹, Isobel Stewart⁹, Carol A. Wang⁴⁴, Yujie Wang³⁷, Peitao Wu¹⁸, Weihua Zhang^{45,46}, Tarunveer S. Ahluwalia^{16,17}, Emil VR Appel⁴⁷, Lawrence F. Bielak⁴⁸, Jennifer A. Brody⁴⁹, Noel P Burt⁵⁰, Claudia P Cabrera^{51,52}, Brian E Cade^{53,54}, Jin Fang Chai⁵⁵, Xiaoran Chai^{56,57}, Li-Ching Chang⁵⁸, Chien-Hsiun Chen⁵⁸, Brian H Chen⁵⁹, Kumaraswamy N Chitrala⁶⁰, Yen-Feng Chiu⁶¹, Hugoline G. de Haan¹⁹, Graciela E Delgado³⁴, Ayse Demirkan⁶², Qing Duan^{3,63}, Jorgen Engmann⁶⁴, Segun A Fatumo^{65,66,67}, Javier Gayán⁶⁸, Franco Giulianini⁶⁹, Jung Ho Gong²⁰, Stefan Gustafsson⁷⁰, Yang Hai⁷¹, Fernando P Hartwig^{72,73}, Jing He⁷⁴, Yoriko Heianza⁷⁵, Tao Huang⁷⁶, Alicia Huerta-Chagoya^{77,78}, Mi Yeong Hwang⁷⁹, Richard A. Jensen⁴⁹, Takahisa Kawaguchi⁸⁰, Katherine A Kentistou^{81,82}, Young Jin Kim⁷⁹, Marcus E Kleber³⁴, Ishminder K Kooner⁸³, Shuiqing Lai⁸⁴, Leslie A Lange⁸⁵, Carl D Langefeld⁸⁶, Marie Lauzon²³, Man Li⁸⁷, Symen Ligthart⁶², Jun Liu^{62,88}, Marie Loh^{89,45}, Jirong Long⁷⁴, Valeriya Lyssenko^{90,91}, Massimo Mangino^{92,93}, Carola Marzi^{94,95}, May E Montasser⁹⁶, Abhishek Nag¹², Masahiro Nakatochi⁹⁷, Damia Noce⁹⁸, Raymond Noordam⁹⁹, Giorgio Pistis¹⁰⁰, Michael Preuss^{39,101}, Laura Raffield³, Laura J. Rasmussen-Torvik¹⁰², Stephen S Rich^{103,104}, Neil R Robertson¹², Rico Rueedi^{105,106}, Kathleen Ryan⁹⁶, Serena Sanna^{100,26}, Richa Saxena^{107,108,109}, Katharina E Schraut^{81,82}, Bengt Sennblad¹¹⁰, Kazuya Setoh⁸⁰, Albert V Smith^{111,112}, Lorraine Southam^{113,114}, Thomas Sparsø⁴⁷, Rona J Strawbridge^{115,116}, Fumihiko Takeuchi¹¹⁷, Jingyi Tan²³, Stella Trompet^{99,118}, Erik van den Akker^{119,120,121}, Peter J Van der Most³⁶, Niek Verweij^{122,123}, Mandy Vogel¹²⁴, Heming Wang⁵³, Chaolong Wang^{125,126}, Nan Wang^{127,128}, Helen R Warren^{51,52}, Wanqing Wen⁷⁴, Tom Wilsgaard¹²⁹, Andrew Wong¹³⁰, Andrew R Wood¹³¹, Tian Xie³⁶, Mohammad Hadi Zafarmand^{132,133}, Jing-Hua Zhao¹³⁴, Wei Zhao⁴⁸, Najaf Amin⁶², Zorayr Arzumanyan²³, Arne Astrup¹³⁵, Stephan JL Bakker¹³⁶, Damiano Baldassarre^{137,138}, Marian Beekman¹¹⁹, Richard N Bergman¹³⁹, Alain Bertoni¹⁴⁰, Matthias Blüher¹⁴¹, Lori L. Bonnycastle¹⁴², Stefan R Bornstein¹⁴³, Donald W Bowden¹⁴⁴, Qiuyin Cai⁷⁴, Archie Campbell^{145,146}, Harry Campbell⁸¹, Yi Cheng Chang^{147,148,149}, Eco J.C. de Geus^{27,28}, Abbas Dehghan⁶², Shufa Du¹⁵⁰, Gudny Eiriksdottir¹¹², Alike Eleni Farmaki^{151,152}, Mattias Frånberg¹⁵³, Christian Fuchsberger⁹⁸, Yutang Gao¹⁵⁴, Anette P Gjesing⁴⁷, Anuj Goel^{155,12}, Sohee Han⁷⁹, Catharina A Hartman¹⁵⁶, Christian Herder^{157,158,159}, Andrew A. Hicks⁹⁸, Chang-Hsun Hsieh^{160,161}, Willa A. Hsueh¹⁶², Sahoko Ichihara¹⁶³, Michiya Igase¹⁶⁴, M. Arfan Ikram⁶², W. Craig Johnson³¹, Marit E Jørgensen^{17,165}, Peter K Joshi⁸¹, Rita R Kalyani¹⁶⁶, Fouad R. Kandeel¹⁶⁷, Tomohiro Katsuya^{168,169}, Chiea Chuen Khor¹²⁶, Wieland Kiess¹²⁴, Ivana Kolcic¹⁷⁰, Teemu Kuulasmaa¹⁷¹, Johanna Kuusisto¹⁷², Kristi Läll¹⁵, Kelvin Lam²³, Deborah A Lawlor^{7,8}, Nanette R. Lee^{173,174}, Rozenn N. Lemaitre⁴⁹, Honglan Li¹⁷⁵, Lifelines Cohort Study¹⁷⁶, Shih-Yi Lin^{177,178,179}, Jaana Lindström¹⁸⁰, Allan Linneberg^{181,182}, Jianjun Liu^{126,183}, Carlos Lorenzo¹⁸⁴, Tatsuaki Matsubara¹⁸⁵, Fumihiko Matsuda⁸⁰, Geltrude Mingrone¹⁸⁶, Simon Mooijaart⁹⁹, Sanghoon Moon⁷⁹, Toru Nabika¹⁸⁷, Girish N. Nadkarni³⁹, Jerry L. Nadler¹⁸⁸, Mari Nelis¹⁵, Matthew J Neville¹¹, Jill M Norris¹⁸⁹, Yasumasa Ohya¹⁹⁰, Annette Peters^{191,95,192}, Patricia A. Peyser⁴⁸, Ozren Polasek¹⁷⁰, Qibin Qi¹⁹³, Dennis Raven¹⁵⁶, Dermot F Reilly¹⁹⁴, Alex Reiner¹⁹⁵, Fernando Rivideneira¹⁹⁶, Kathryn Roll²³, Igor Rudan¹⁹⁷, Charumathi Sabanayagam^{56,198}, Kevin Sandow²³, Naveed Sattar¹⁹⁹, Annette Schürmann^{200,201}, Jinxiu Shi²⁰², Heather M Stringham^{43,203}, Kent D. Taylor²³, Tanya M. Teslovich²⁰⁴, Betina Thuesen¹⁸¹, Paul RHJ Timmers⁸¹, Elena Tremoli¹³⁸, Michael Y Tsai²⁰⁵, Andre Uitterlinden¹⁹⁶, Rob M van Dam^{41,183,206}, Diana van Heemst⁹⁹, Astrid van Hylckama Vlieg¹⁹, Jana V Van Vliet-Ostaptchouk³⁶, Jagadish Vangipurapu²⁰⁷, Henrik Vestergaard^{47,17}, Tao Wang¹⁹³, Ko Willems van Dijk^{208,209,210}, Yongbing Xiang¹⁷⁵, Tatijana Zemunik²¹¹, Goncalo R Abecasis⁴³, Linda S. Adair^{150,212}, Carlos Alberto Aguilar-Salinas^{213,214,215}, Marta E Alarcón-Riquelme^{216,217}, Ping An²¹⁸, Larissa Aviles-Santa²¹⁹,

Diane M Becker²²⁰, Lawrence J Beilin²²¹, Sven Bergmann^{105,106,222}, Hans Bisgaard^{16,17}, Corri Black²²³, Michael Boehnke^{224,225}, Eric Boerwinkle^{226,227}, Bernhard O Böhm^{228,229}, Klaus Bønnelykke^{16,17}, D I. Boomsma^{27,28}, Erwin P. Bottinger^{39,40,230}, Thomas A Buchanan^{231,232,128}, Mickaël Canouil^{233,234}, Mark J Caulfield^{51,52}, John C. Chambers^{89,45,83,235,236}, Daniel I. Chasman^{69,237}, Yii-Der Ida Chen²³, Ching-Yu Cheng^{56,198}, Francis S. Collins¹⁴², Adolfo Correa²³⁸, Francesco Cucca¹⁰⁰, H. Janaka de Silva²³⁹, George Dedoussis²⁴⁰, Sölve Elmståhl²⁴¹, Michele K. Evans⁶⁰, Ele Ferranni²⁴², Luigi Ferruci²⁴³, Jose C Florez^{244,245,246}, Paul Franks²⁴⁷, Timothy M Frayling¹³¹, Philippe Froguel^{233,234,248}, Bruna Gigante²⁴⁹, Mark O. Goodarzi²⁵⁰, Penny Gordon-Larsen^{150,212}, Harald Grallert^{94,95}, Niels Grarup⁴⁷, Sameline Grimsgaard¹²⁹, Leif Groop^{251,252}, Vilundur Gudnason^{112,253}, Xiuqing Guo²³, Anders Hamsten¹¹⁶, Torben Hansen⁴⁷, Caroline Hayward²⁵⁴, Susan R. Heckbert²⁵⁵, Bernardo L Horta⁷², Wei Huang²⁰², Erik Ingelsson²⁵⁶, Pankow S James²⁵⁷, Jost B Jonas^{258,259}, J. Wouter Jukema^{118,260}, Pontiano Kaleebu²⁶¹, Robert Kaplan^{193,195}, Sharon L.R. Kardia⁴⁸, Norihiro Kato¹¹⁷, Sirkka M. Keinänen-Kiukaanniemi^{262,263}, Bong-Jo Kim⁷⁹, Mika Kivimäki²⁶⁴, Heikki A. Koistinen^{265,266,267}, Jaspal S. Kooner^{83,235,236,268}, Antje Körner¹²⁴, Peter Kovacs^{141,269}, Diana Kuh¹³⁰, Meena Kumari²⁷⁰, Zoltan Kutalik^{271,106}, Markku Laakso¹⁷², Timo A. Lakka^{272,273,274}, Lenore J Launer²⁷⁵, Karin Leander²⁷⁶, Huaixing Li²⁹, Xu Lin²⁹, Lars Lind²⁷⁷, Cecilia Lindgren^{12,278,279}, Simin Liu⁸⁴, Ruth J.F. Loos^{39,101}, Patrik Magnusson²⁸⁰, Anubha Mahajan¹², Andres Metspalu¹⁵, Dennis O Mook-Kanamori^{19,281}, Trevor A Mori²²¹, Patricia B Munroe^{51,52}, Inger Njølstad¹²⁹, Jeffrey R O'Connell⁹⁶, Albertine J Oldehinkel¹⁵⁶, Ken K Ong⁹, Sandosh Padmanabhan²⁸², Colin N.A. Palmer²⁸³, Nicholette D Palmer¹⁴⁴, Oluf Pedersen⁴⁷, Craig E Pennell⁴⁴, David J Porteous^{145,284}, Peter P. Pramstaller⁹⁸, Michael A. Province²¹⁸, Bruce M. Psaty^{49,255,285,286}, Lu Qi⁷⁵, Leslie J. Raffel²⁸⁷, Rainer Rauramaa²⁷⁴, Susan Redline^{53,54}, Paul M Ridker^{69,237}, Frits R. Rosendaal¹⁹, Timo E. Saaristo^{288,289}, Manjinder Sandhu²⁹⁰, Jouko Saramies²⁹¹, Neil Schneiderman²⁹², Peter Schwarz^{143,293,201}, Laura J. Scott^{224,225}, Elizabeth Selvin³⁸, Peter Sever²⁶⁸, Xiao-ou Shu⁷⁴, P Eline Slagboom¹¹⁹, Kerrin S Small⁹², Blair H Smith²⁹⁴, Harold Snieder³⁶, Tamar Sofer^{295,245}, Thorkild I.A. Sørensen^{296,297,7,8}, Tim D Spector⁹², Alice Stanton²⁹⁸, Claire J Steves^{92,299}, Michael Stumvoll¹⁴¹, Liang Sun²⁹, Yasuharu Tabara⁸⁰, E Shyong Tai^{183,300,301}, Nicholas J Timpson^{7,8}, Anke Tönjes¹⁴¹, Jaakko Tuomilehto^{302,303,304}, Teresa Tusie^{78,305}, Matti Uusitupa³⁰⁶, Pim van der Harst^{122,26}, Cornelia van Duijn^{307,62}, Veronique Vitart²⁵⁴, Peter Vollenweider³⁰⁸, Tanja GM Vrijkotte¹³², Lynne E Wagenknecht³⁰⁹, Mark Walker³¹⁰, Ya X Wang²⁵⁹, Nick J Wareham³¹¹, Richard M Watanabe^{127,232,128}, Hugh Watkins^{155,12}, Wen B Wei³¹², Ananda R Wickremasinghe³¹³, Gonneke Willemsen^{27,28}, James F Wilson^{81,254}, Tien-Yin Wong^{56,198}, Jer-Yuarn Wu⁵⁸, Anny H Xiang³¹⁴, Lisa R Yanek²²⁰, Loïc Yengo^{233,234,315}, Mitsuhiro Yokota³¹⁶, Eleftheria Zeggini^{317,318,319}, Wei Zheng³²⁰, Alan B Zonderman⁶⁰, Jerome I Rotter²³, Anna L Gloyn^{11,12,321,322}, Mark I. McCarthy^{11,323,324}, Josée Dupuis¹⁸, James B Meigs^{325,245,326}, Robert Scott⁹, Inga Prokopenko^{30,24}, Aaron Leong^{327,328,329}, Ching-Ti Liu¹⁸, Stephen CJ Parker^{6,330*}, Karen L. Mohlke^{3*}, Claudia Langenberg^{9*}, Eleanor Wheeler^{9,1*}, Andrew P. Morris^{331,332,12*}, Inês Barroso^{1,333,2*}

¹Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, UK, ²Exeter Centre of Excellence in Diabetes (ExCEeD), Exeter Medical School, University of Exeter, Exeter, UK, ³Department of Genetics, University of North Carolina, Chapel Hill, NC, USA, ⁴Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA, USA, ⁵Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France, ⁶Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, ⁷MRC Integrative Epidemiology Unit, University of Bristol, Bristol, Bristol, UK, ⁸Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, Bristol, UK, ⁹MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK, ¹⁰Department of Biostatistics, School of Public Health, Shandong University, Jinan, Shandong, China, ¹¹Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK, ¹²Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK, ¹³Laboratory for Genomics of Diabetes and Metabolism, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan, ¹⁴Medical Research Council Human Genetics Unit, Institute for Genetics and Molecular Medicine, Edinburgh, UK, ¹⁵Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia, ¹⁶COPSAC,

Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark, ¹⁷Steno Diabetes Center Copenhagen, Gentofte, Denmark, ¹⁸Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, ¹⁹Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, ²⁰Department of Epidemiology, Brown University, Providence, RI, USA, ²¹Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China, ²²Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China, ²³The Institute for Translational Genomics and Population Sciences, ` , The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA, ²⁴Department of Metabolism, Digestion, and Reproduction, Imperial College London, London, UK, ²⁵Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA, ²⁶Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, ²⁷Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, ²⁸Amsterdam Public Health Research Institute, Amsterdam Universities Medical Center, Amsterdam, The Netherlands, ²⁹CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, ³⁰Department of Clinical and Experimental Medicine, Section of Statistical Multi-Omics, University of Surrey, Guildford, UK, ³¹Department of Biostatistics, University of Washington, Seattle, WA, USA, ³²SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Mannheim, Germany, ³³Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, Graz, Austria, ³⁴Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, Germany, ³⁵Department of Economics, Metropolitan Autonomous University, Mexico City, Mexico, ³⁶Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, ³⁷CVD Genetic Epidemiology Computational Laboratory, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, ³⁸Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA, ³⁹The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, ⁴⁰HPI Digital Health Center, Digital Health and Personalized Medicine, Hasso Plattner Institute, Potsdam, Germany, ⁴¹Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, ⁴²Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, ⁴³Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA, ⁴⁴School of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Newcastle, NSW, Australia, ⁴⁵Department of Epidemiology and Biostatistics, Imperial College London, London, UK, ⁴⁶Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UK, ⁴⁷Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ⁴⁸Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA, ⁴⁹Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA, ⁵⁰Metabolism Program, Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, ⁵¹Department of Clinical Pharmacology, William Harvey Research Institute Barts, The London School of Medicine and Dentistry,, Queen Mary University of London, London, UK, ⁵²NIHR Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK, ⁵³Department of Medicine, Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, ⁵⁴Department of Medicine, Sleep Medicine, Harvard Medical School, Boston, MA, USA, ⁵⁵Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, ⁵⁶Ocular Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore, ⁵⁷Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore, ⁵⁸Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, Taiwan, ⁵⁹Department of Epidemiology, UCLA, Los Angeles, CA, USA, ⁶⁰Laboratory of Epidemiology and Population Sciences,

154 National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore,
155 MD, USA, ⁶¹Institute of Population Health Sciences, National Health Research Institutes, Miaoli,
156 Taiwan, ⁶²Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands,
157 ⁶³Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA,
158 ⁶⁴Institute of Cardiovascular Science, UCL, London, UK, ⁶⁵Uganda Medical Informatics Centre (UMIC),
159 MRC/UVRI and London School of Hygiene & Tropical Medicine (Uganda Research Unit), Entebbe,
160 Uganda, ⁶⁶London School of Hygiene & Tropical Medicine, London, UK, ⁶⁷H3Africa Bioinformatics
161 Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja,
162 Nigeria, ⁶⁸Bioinfosol, Sevilla, Spain, ⁶⁹Division of Preventive Medicine, Brigham and Women's
163 Hospital, Boston, MA, USA, ⁷⁰Department of Medical Sciences, Molecular Epidemiology and Science
164 for Life Laboratory, Uppsala University, Uppsala, Sweden, ⁷¹Department of Statistics, The University
165 of Auckland, Science Center, Auckland, New Zealand, ⁷²Postgraduate Program in Epidemiology,
166 Federal University of Pelotas, Pelotas, RS, Brazil, ⁷³MRC Integrative Epidemiology Unit, University of
167 Bristol, Bristol, UK, ⁷⁴Department of Medicine, Epidemiology, Vanderbilt University Medical Center,
168 Nashville, TN, USA, ⁷⁵Department of Epidemiology, Tulane University Obesity Research Center,,
169 Tulane University, New Orleans, USA, ⁷⁶Department of Epidemiology and Biostatistics, School of
170 Public Health, Peking University, Beijing, China, ⁷⁷Molecular Biology and Genomic Medicine Unit,
171 National Council for Science and Technology, Mexico City, Mexico, ⁷⁸Molecular Biology and Genomic
172 Medicine Unit, National Institute of Medical Sciences and Nutrition, Mexico City, Mexico, ⁷⁹Division
173 of Genome Research, Center for Genome Research, Korea National Institute of Health, Cheongju-si,
174 Chungcheongbuk-do, South Korea, ⁸⁰Center for Genomic Medicine, Kyoto University Graduate
175 School of Medicine, Kyoto, Japan, ⁸¹Centre for Global Health Research, Usher Institute of Population
176 Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, ⁸²Centre for
177 Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh,
178 Scotland, ⁸³Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust,
179 Middlesex, UK, ⁸⁴Department of Epidemiology, Brown University School of Public Health, Providence,
180 RI, USA, ⁸⁵Department of Medicine, Division of Biomedical Informatics and Personalized Medicine,
181 University of Colorado Anschutz Medical Campus, Denver, CO, USA, ⁸⁶Department of Public Health
182 Sciences, Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA,
183 ⁸⁷Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake
184 City, UT, USA, ⁸⁸Nuffield Department of Population Health, University of Oxford, Oxford, UK, ⁸⁹Lee
185 Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore,
186 ⁹⁰Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen,
187 Norway, ⁹¹Department of Clinical Sciences, Lund University Diabetes Centre, Lund University,
188 Malmo, Sweden, ⁹²Department of Twin Research and Genetic Epidemiology, School of Life Course
189 Sciences, King's College London, London, UK, ⁹³NIHR Biomedical Research Centre, Guy's and St
190 Thomas' Foundation Trust, London, UK, ⁹⁴Institute of Epidemiology, Research Unit of Molecular
191 Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg,
192 Bavaria, Germany, ⁹⁵German Center for Diabetes Research (DZD), Neuherberg, Bavaria, Germany,
193 ⁹⁶Department of Medicine, Division of Endocrinology, Diabetes, and Nutrition, University of
194 Maryland School of Medicine, Baltimore, MD, USA, ⁹⁷Public Health Informatics Unit, Department of
195 Integrated Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan, ⁹⁸Institute for
196 Biomedicine, Eurac Research, Bolzano, BZ, Italy, ⁹⁹Department of Internal Medicine, Section of
197 Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands, ¹⁰⁰Istituto di
198 Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche (CNR), Monserrato, Italy,
199 Italy, ¹⁰¹The Mindich Child Health and Development Institute for Personalized Medicine, Icahn School
200 of Medicine at Mount Sinai, New York, NY, USA, ¹⁰²Department of Preventive Medicine,
201 Northwestern University Feinberg School of Medicine, Chicago, IL, USA, ¹⁰³Center for Public Health
202 Genomics, University of Virginia, Charlottesville, VA, USA, ¹⁰⁴Department of Public Health Sciences,
203 University of Virginia, Charlottesville, VA, USA, ¹⁰⁵Department of Computational Biology, University
204 of Lausanne, Lausanne, Switzerland, ¹⁰⁶Swiss Institute of Bioinformatics, Lausanne, Switzerland,

205 ¹⁰⁷Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston,
 206 MA, USA, ¹⁰⁸Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General
 207 Hospital, Boston, MA, USA, ¹⁰⁹Program in Medical and Population Genetics,, Broad Institute,
 208 Cambridge, MA, USA, ¹¹⁰Department of Cell and Molecular Biology., National Bioinformatics
 209 Infrastructure Sweden,, Science for Life Laboratory, Uppsala University, Uppsala, Sweden,
 210 ¹¹¹Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA, ¹¹²Icelandic Heart
 211 Association, Kopavogur, Iceland, ¹¹³Institute of Translational Genomics, Helmholtz Zentrum
 212 München – German Research Center for Environmental Health, Institute of Translational Genomics,
 213 Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg,
 214 Germany, ¹¹⁴Wellcome Sanger Institute, Hinxton, Cambridge, UK, ¹¹⁵Institute of Health and
 215 Wellbeing, University of Glasgow, Glasgow, Glasgow, UK, ¹¹⁶Department of Medicine Solna,
 216 Cardiovascular medicine, Karolinska Institutet, Stockholm, Sweden, ¹¹⁷National Center for Global
 217 Health and Medicine, Tokyo, Japan, ¹¹⁸Department of Cardiology, Leiden University Medical Center,
 218 Leiden, The Netherlands, ¹¹⁹Department of Biomedical Data Sciences, Molecular Epidemiology,
 219 Leiden University Medical Center, Leiden, The Netherlands, ¹²⁰Department of Pattern Recognition &
 220 Bioinformatics, Delft University of Technology, Delft, The Netherlands, ¹²¹Department of Biomedical
 221 Data Sciences, Leiden Computational Biology Center, Leiden University Medical Center, Leiden, The
 222 Netherlands, ¹²²Department of Cardiology, University Medical Center Groningen, University of
 223 Groningen, Groningen, The Netherlands, ¹²³Genomics plc, Oxford, UK, ¹²⁴Center of Pediatric
 224 Research, University Children’s Hospital Leipzig, University of Leipzig Medical Center, Leipzig,
 225 Germany, ¹²⁵Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical
 226 College, Huazhong University of Science and Technology, Wuhan, China, ¹²⁶Genome Institute of
 227 Singapore, Agency for Science, Technology and Research, Singapore, Singapore, ¹²⁷Department of
 228 Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA, ¹²⁸USC Diabetes and
 229 Obesity Research Institute, Keck School of Medicine of USC, Los Angeles, CA, USA, ¹²⁹Department of
 230 Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø,
 231 Norway, ¹³⁰MRC Unit for Lifelong Health & Ageing at UCL, London, UK, ¹³¹Department of Human
 232 Genetics, University of Exeter RILD Building Royal Devon and Exeter NHS trust Barrack Road, Exeter,
 233 UK, ¹³²Department of Public Health, Amsterdam Public Health Research Institute, Amsterdam
 234 Universities Medical Center, Amsterdam, The Netherlands, ¹³³Department of Clinical Epidemiology,
 235 Biostatistics, and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam
 236 Universities Medical Center, Amsterdam, The Netherlands, ¹³⁴Department of Public Health and
 237 Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK, ¹³⁵Department of
 238 Nutrition, Exercise, and Sports, Faculty of Science, University of Copenhagen, Copenhagen,
 239 Denmark, ¹³⁶Department of Internal Medicine, University Medical Center Groningen, University of
 240 Groningen, Groningen, The Netherlands, ¹³⁷Department of Medical Biotechnology and Translational
 241 Medicine, University of Milan, Milan, Milan, Italy, ¹³⁸Centro Cardiologico Monzino, IRCCS, Milan,
 242 Italy, ¹³⁹Department of Physiology and Biophysics, University of Southern California, Los Angeles, CA,
 243 USA, ¹⁴⁰Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest
 244 School of Medicine, Winston-Salem, NC, USA, ¹⁴¹Medical Department III – Endocrinology,
 245 Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany, ¹⁴²Medical
 246 Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National
 247 Institutes of Health, Bethesda, MD, USA, ¹⁴³Department for Prevention and Care of Diabetes, Faculty
 248 of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, ¹⁴⁴Department
 249 of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA, ¹⁴⁵Centre for Genomic
 250 and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh,
 251 Western General Hospital, Edinburgh, UK, ¹⁴⁶Usher Institute for Population Health Sciences and
 252 Informatics, University of Edinburgh, Edinburgh, UK, ¹⁴⁷Internal Medicine, National Taiwan University
 253 Hospital, Taipei, Taiwan, ¹⁴⁸Medical Genomics and Proteomics, National Taiwan University, Taipei,
 254 Taiwan, ¹⁴⁹Biomedical Sciences, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan,
 255 ¹⁵⁰Department of Nutrition, Gillings School of Global Public Health, University of North Carolina,

256 Chapel Hill, NC, USA, ¹⁵¹Department of Population Science and Experimental Medicine, Institute of
 257 Cardiovascular Science, University College London, London, UK, ¹⁵²Department of Nutrition and
 258 Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece,
 259 ¹⁵³Department of Medicine Solna, Cardiovascular medicine, Stockholm, Sweden, ¹⁵⁴Department of
 260 Epidemiology, Shanghai Cancer Institute, Shanghai, China, ¹⁵⁵Division of Cardiovascular Medicine,
 261 Radcliffe Department of Medicine, University of Oxford, Oxford, UK, ¹⁵⁶Department of Psychiatry,
 262 Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center
 263 Groningen, University of Groningen, Groningen, The Netherlands, ¹⁵⁷Institute for Clinical
 264 Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine
 265 University Düsseldorf, Düsseldorf, Germany, ¹⁵⁸Division of Endocrinology and Diabetology, Medical
 266 Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, ¹⁵⁹German Center for Diabetes
 267 Research (DZD), Düsseldorf, Germany, ¹⁶⁰Internal Medicine, Endocrine & Metabolism, Tri-Service
 268 General Hospital, Taipei, Taiwan, ¹⁶¹School of Medicine, National Defense Medical Center, Taipei,
 269 Taiwan, ¹⁶²Internal Medicine, Endocrinology, Diabetes & Metabolism, Diabetes and Metabolism
 270 Research Center, The Ohio State University Wexner Medical Center, Columbus, OH, USA,
 271 ¹⁶³Department of Environmental and Preventive Medicine, Jichi Medical University School of
 272 Medicine, Shimotsuke, Japan, ¹⁶⁴Department of Anti-aging Medicine, Ehime University Graduate
 273 School of Medicine, Toon, Japan, ¹⁶⁵National Institute of Public Health, University of Southern
 274 Denmark, Odense, Denmark, ¹⁶⁶Department of Medicine, Endocrinology, Diabetes & Metabolism,
 275 Johns Hopkins University School of Medicine, Baltimore, MD, USA, ¹⁶⁷Clinical Diabetes,
 276 Endocrinology & Metabolism, Translational Research & Cellular Therapeutics, Beckman Research
 277 Institute of the City of Hope, Duarte, CA, USA, ¹⁶⁸Department of Clinical Gene Therapy, Osaka
 278 University Graduate School of Medicine, Suita, Japan, ¹⁶⁹Department of Geriatric and General
 279 Medicine, Osaka University Graduate School of Medicine, Suita, Japan, ¹⁷⁰Department of Public
 280 Health, University of Split School of Medicine, Split, Croatia, ¹⁷¹Institute of Biomedicine,
 281 Bioinformatics Center, University of Eastern Finland, Kuopio, Finland, ¹⁷²Department of Medicine,
 282 University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland, ¹⁷³USC-Office of
 283 Population Studies Foundation, University of San Carlos, Cebu City, Philippines, ¹⁷⁴Department of
 284 Anthropology, Sociology and History, University of San Carlos, Cebu City, Philippines, ¹⁷⁵State Key
 285 Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer
 286 Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China,
 287 ¹⁷⁶Lifelines cohort, Groningen, The Netherlands, ¹⁷⁷Internal Medicine, Endocrine & Metabolism,
 288 Taichung Veterans General Hospital, Taichung, Taiwan, ¹⁷⁸Center for Geriatrics and Gerontology,,
 289 Taichung Veterans General Hospital, Taichung, Taiwan, ¹⁷⁹National Defense Medical Center, National
 290 Yang-Ming University, Taipei, Taiwan, ¹⁸⁰Diabetes Prevention Unit, National Institute for Health and
 291 Welfare, Helsinki, Finland, ¹⁸¹Center for Clinical Research and Prevention, Bispebjerg and
 292 Frederiksberg Hospital, Copenhagen, Denmark, ¹⁸²Department of Clinical Medicine, Faculty of Health
 293 and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ¹⁸³Yong Loo Lin School of
 294 Medicine, National University of Singapore and National University Health System, Singapore,
 295 Singapore, ¹⁸⁴Department of Medicine, University of Texas Health Sciences Center, San Antonio, TX,
 296 USA, ¹⁸⁵Department of Internal Medicine, Aichi Gakuin University School of Dentistry, Nagoya, Japan,
 297 ¹⁸⁶Department of Diabetes, Diabetes, & Nutritional Sciences, James Black Centre, King's College
 298 London, London, UK, ¹⁸⁷Department of Functional Pathology, Shimane University School of
 299 Medicine, Izumo, Japan, ¹⁸⁸Department of Medicine and Pharmacology, New York Medical College
 300 School of Medicine, Valhalla, NY, USA, ¹⁸⁹Colorado School of Public Health, University of Colorado
 301 Anschutz Medical Campus, Aurora, CO, USA, ¹⁹⁰Department of Geriatric Medicine and Neurology,
 302 Ehime University Graduate School of Medicine, Toon, Japan, ¹⁹¹Institute of Epidemiology, Helmholtz
 303 Zentrum München Research Center for Environmental Health, Neuherberg, Bavaria, Germany,
 304 ¹⁹²Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians
 305 University Munich, Munich, Bavaria, Germany, ¹⁹³Department of Epidemiology and Population
 306 Health, Albert Einstein College of Medicine, Bronx, NY, USA, ¹⁹⁴Genetics and Pharmacogenomics,

307 Merck Sharp & Dohme Corp., Kenilworth, NJ, USA,¹⁹⁵Department of Public Health Sciences, Fred
308 Hutchinson Cancer Research Center, Seattle, WA, USA,¹⁹⁶Department of Internal Medicine, Erasmus
309 Medical Center, Rotterdam, The Netherlands,¹⁹⁷Usher Institute of Population Health Sciences and
310 Informatics, University of Edinburgh, Edinburgh, UK,¹⁹⁸Ophthalmology & Visual Sciences Academic
311 Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore,¹⁹⁹BHF Glasgow
312 Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of
313 Glasgow, Glasgow, UK,²⁰⁰Department of Experimental Diabetology, German Institute of Human
314 Nutrition Potsdam-Rehbruecke, Nuthetal, Germany,²⁰¹German Center for Diabetes Research (DZD),
315 Neuherberg, Germany,²⁰²Department of Genetics, Shanghai-MOST Key Laboratory of Health and
316 Disease Genomics, Chinese National Human Genome Center at Shanghai (CHGC) and Shanghai
317 Academy of Science & Technology (SAST), Shanghai, China,²⁰³Center for Statistical Genetics, Ann
318 Arbor, MI, USA,²⁰⁴Sarepta Therapeutics, Cambridge, Massachusetts, USA,²⁰⁵Department of
319 Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA,²⁰⁶Department
320 of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA,²⁰⁷Institute of Clinical
321 Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland,²⁰⁸Department of
322 Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The
323 Netherlands,²⁰⁹Laboratory for Experimental Vascular Medicine, Leiden University Medical Center,
324 Leiden, The Netherlands,²¹⁰Department of Human Genetics, Leiden University Medical Center,
325 Leiden, The Netherlands,²¹¹Department of Human Biology, University of Split School of Medicine,
326 Split, Croatia,²¹²Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA,
327²¹³Department of Endocrinology and Metabolism, Instituto Nacional de Ciencias Medicas y Nutricion,
328 Mexico City, Mexico,²¹⁴Unidad de Investigacion de Enfermedades Metabolicas. Tec Salud,
329 Mexico City, Mexico,²¹⁵Instituto Tecnológico y de Estudios Superiores de Monterrey Tec Salud,
330 Mexico City, Mexico,²¹⁶Department of Medical Genomics, Pfizer/University of Granada/Andalusian
331 Government Center for Genomics and Oncological Research (GENYO), Granada, Spain,²¹⁷Institute
332 for Environmental Medicine, Chronic Inflammatory Diseases, Karolinska Institutet, Solna, Sweden,
333²¹⁸Department of Genetics, Division of Statistical Genomics, Washington University School of
334 Medicine, St. Louis, MO, USA,²¹⁹Clinical and Health Services Research, National Institute on Minority
335 Health and Health Disparities, Bethesda, MD, USA,²²⁰Department of Medicine, General Internal
336 Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA,²²¹School of Medicine
337 and Pharmacology, Royal Perth Hospital Unit, University of Western Australia, Perth, WA, Australia,
338²²²Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South
339 Africa,²²³Aberdeen Centre for Health Data Science, 1:042 Polwarth Building,, School of Medicine,
340 Medical, Science and Nutrition, University of Aberdeen, Foresterhill, Aberdeen, UK,²²⁴Department of
341 Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA,²²⁵Center for
342 Statistical Genetics, University of Michigan, Ann Arbor, MI, USA,²²⁶Human Genetics Center, School
343 of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA,
344²²⁷Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA,²²⁸Division of
345 Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University
346 of Ulm, Ulm, Baden-Württemberg, Germany,²²⁹LKC School of Medicine, Nanyang Technological
347 University, Singapore and Imperial College London, UK, Singapore, Singapore,²³⁰University of
348 Potsdam, Berlin-Potsdam, Germany,²³¹Department of Medicine, Keck School of Medicine of USC,
349 Los Angeles, CA, USA,²³²Department of Physiology and Neuroscience, Keck School of Medicine of
350 USC, Los Angeles, CA, USA,²³³INSERM UMR 1283 / CNRS UMR 8199, European Institute for Diabetes
351 (EGID), Université de Lille, Lille, France,²³⁴INSERM UMR 1283 / CNRS UMR 8199, European Institute
352 for Diabetes (EGID), Institut Pasteur de Lille, Lille, France,²³⁵Imperial College Healthcare NHS Trust,
353 Imperial College London, London, UK,²³⁶MRC-PHE Centre for Environment and Health, Imperial
354 College London, London, UK,²³⁷Havard Medical School, Boston, MA, USA,²³⁸Department of
355 Medicine, Jackson Heart Study, University of Mississippi Medical Center, Jackson, MS, USA,
356²³⁹Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka,
357²⁴⁰Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio

358 University of Athens, Kallithea, Greece, ²⁴¹Department of Clinical Sciences, Lund University, Malmö,
 359 Sweden, ²⁴²CNR Institute of Clinical Physiology, Pisa, Italy, ²⁴³Intramural Research Program, National
 360 Institute of Aging, Baltimore, MD, USA, ²⁴⁴Diabetes Unit and Center for Genomic Medicine,
 361 Massachusetts General Hospital, Boston, MA, USA, ²⁴⁵Department of Medicine, Harvard Medical
 362 School, Boston, MA, USA, ²⁴⁶Programs in Metabolism and Medical & Population Genetics, Broad
 363 Institute, Cambridge, MA, USA, ²⁴⁷Department of Clinical Sciences, Lund University Diabetes Centre,
 364 Malmö, Sweden, ²⁴⁸Department of Genomics of Common Disease, Imperial College London,
 365 London, UK, ²⁴⁹Department of Medicine, Cardiovascular medicine, Karolinska Institutet, Stockholm,
 366 Sweden, ²⁵⁰Department of Medicine, Division of Endocrinology, Diabetes & Metabolism, Cedars-Sinai
 367 Medical Center, Los Angeles, CA, USA, ²⁵¹Diabetes Centre, Lund University, Sweden, ²⁵²Finnish
 368 Institute of Molecular Medicine, Helsinki University, Helsinki, Finland, ²⁵³Faculty of Medicine, School
 369 of health sciences, University of Iceland, Reykjavik, Iceland, ²⁵⁴MRC Human Genetics Unit, Institute of
 370 Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, ²⁵⁵Department of
 371 Epidemiology, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA,
 372 ²⁵⁶Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of
 373 Medicine, Stanford University, Stanford, CA, USA, ²⁵⁷Division of Epidemiology and Community
 374 Health, University of Minnesota, Minneapolis, MN, USA, ²⁵⁸Department of Ophthalmology, Medical
 375 Faculty Mannheim, Heidelberg University, Mannheim, Germany, ²⁵⁹Beijing Institute of
 376 Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center,
 377 Beijing Tongren Hospital, Capital Medical University, Beijing, China, ²⁶⁰Netherlands Heart Institute,
 378 Utrecht, The Netherlands, ²⁶¹MRC/UVRI and LSHTM (Uganda Research Unit), Entebbe, Uganda,
 379 ²⁶²Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland, ²⁶³Unit of
 380 General Practice, Oulu University Hospital, Oulu, Finland, ²⁶⁴Department of Epidemiology and Public
 381 Health, UCL, London, UK, ²⁶⁵Department of Public Health Solutions, Finnish Institute for Health and
 382 Welfare, Helsinki, Finland, ²⁶⁶Department of Medicine, University of Helsinki and Helsinki University
 383 Central Hospital, Helsinki, Finland, ²⁶⁷Minerva Foundation Institute for Medical Research, Helsinki,
 384 Finland, ²⁶⁸National Heart and Lung Institute, Imperial College London, London, UK, ²⁶⁹IFB Adiposity
 385 Diseases, University of Leipzig Medical Center, Leipzig, Germany, ²⁷⁰Institute for Social and Economic
 386 Research, University of Essex, Colchester, UK, ²⁷¹University Institute of Primary Care and Public
 387 Health, Division of Biostatistics, University of Lausanne, Lausanne, Switzerland, ²⁷²Institute of
 388 Biomedicine, School of Medicine, University of Eastern Finland, Finland, ²⁷³Department of Clinical
 389 Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland, ²⁷⁴Foundation for
 390 Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio,
 391 Finland, ²⁷⁵Laboratory of Epidemiology and Population Sciences, National Institute on Aging,
 392 National Institutes of Health, Baltimore, MD, USA, ²⁷⁶Institute of Environmental Medicine,
 393 Cardiovascular and Nutritional Epidemiology, Karolinska Institutet, Stockholm, Sweden,
 394 ²⁷⁷Department of Medical Sciences, Uppsala, Sweden, ²⁷⁸Big Data Institute, Nuffield Department of
 395 Medicine, University of Oxford, Oxford, UK, ²⁷⁹Nuffield Department of Women's and Reproductive
 396 Health, University of Oxford, Oxford, UK, ²⁸⁰Department of Medical Epidemiology and Biostatistics
 397 and the Swedish Twin Registry, Karolinska Institute, Stockholm, Sweden, ²⁸¹Department of Public
 398 Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands, ²⁸²Institute of
 399 Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK, ²⁸³Division of Population
 400 Health and Genomics, School of Medicine, University of Dundee, Ninewells Hospital and Medical
 401 School, Dundee, UK, ²⁸⁴Centre for Cognitive Ageing and Cognitive Epidemiology, University of
 402 Edinburgh, Edinburgh, UK, ²⁸⁵Department of Health Services, Cardiovascular Health Research Unit,
 403 University of Washington, Seattle, WA, USA, ²⁸⁶Kaiser Permanente Washington Health Research
 404 Institute, Seattle, WA, USA, ²⁸⁷Department of Pediatrics, Genetic and Genomic medicine, University
 405 of California, Irvine, Irvine, CA, USA, ²⁸⁸Tampere, Finnish Diabetes Association, Tampere, Finland,
 406 ²⁸⁹Pirkanmaa Hospital District, Tampere, Finland, ²⁹⁰Department of Medicine, University of
 407 Cambridge, Cambridge, UK, ²⁹¹South Karelia Central Hospital, Lappeenranta, Finland, ²⁹²Department
 408 of Psychology, University of Miami, Miami, FL, USA, ²⁹³Paul Langerhans Institute Dresden of the

Helmholtz Center Munich, University Hospital and Faculty of Medicine, Dresden, Germany,
²⁹⁴Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of
Dundee, Dundee, UK, ²⁹⁵Division of Sleep and Circadian Disorders, Brigham and Women's Hospital,
Boston, MA, USA, ²⁹⁶Novo Nordisk Foundation Center for Basic Metabolic Research, Section of
Metabolic Genetics,, Faculty of Health and Medical Sciences, University of Copenhagen,
Copenhagen, Denmark, ²⁹⁷Department of Public Health, Section of Epidemiology, Faculty of Health
and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ²⁹⁸Department of Molecular
and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland, ²⁹⁹Department of
Aging and Health, Guy's and St Thomas' Foundation Trust, London, UK, ³⁰⁰Saw Swee Hock School of
Public Health, National University of Singapore and National University Health System, Singapore,
Singapore, ³⁰¹Duke-NUS Medical School, Singapore, Singapore, ³⁰²Department of Public Health
Solutions, National Institute for Health and Welfare, Helsinki, Finland, ³⁰³Department of Public
Health, University of Helsinki, Helsinki, Finland, ³⁰⁴Saudi Diabetes Research Group, King Abdulaziz
University, Jeddah, Saudi Arabia, ³⁰⁵Department of Genomic Medicine and Environmental
Toxicology, Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico,
Mexico City, Mexico, ³⁰⁶Department of Public Health and Clinical Nutrition, University of Eastern
Finland, Finland, ³⁰⁷Department of Epidemiology, University of Oxford, Oxford, UK, ³⁰⁸Department of
Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland,
³⁰⁹Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA,
³¹⁰Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK, ³¹¹MRC
Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom, ³¹²Beijing Tongren Eye
Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing
Ophthalmology & Visual Sciences Key Lab, Beijing Tongren Hospital, Capital Medical University,
Beijing, China, China, ³¹³Department of Public Health, Faculty of Medicine, University of Kelaniya,
Ragama, Sri Lanka, ³¹⁴Department of Research and Evaluation, Kaiser Permanente of Southern
California, Pasadena, CA, USA, ³¹⁵Institute for Molecular Bioscience, The University of Queensland,
Queensland, Australia, ³¹⁶Kurume University School of Medicine, Japan, ³¹⁷Institute of Translational
Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health,
Neuherberg, Germany, ³¹⁸Wellcome Sanger Institute, Hinxton, UK, ³¹⁹TUM School of Medicine,
Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany, ³²⁰Department of
Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt University Medical
Center, Nashville, TN, USA, ³²¹NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford,
UK, ³²²Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford,
CA, USA, ³²³Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of
Oxford, Oxford, UK, ³²⁴Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS
Foundation Trust, Oxford, UK, ³²⁵Department of Medicine, Division of General Internal Medicine,
Massachusetts General Hospital, Boston, MA, USA, ³²⁶Program in Medical and Population Genetics,
Broad Institute, Cambridge, MA, USA, ³²⁷Department of Medicine, General Internal Medicine,
Massachusetts General Hospital, Boston, MA, USA, ³²⁸Department of Medicine, Diabetes Unit and
Endocrine Unit, Massachusetts General Hospital, Boston, MA, USA, ³²⁹Harvard Medical School,
Boston, MA, USA, ³³⁰Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA,
³³¹Department of Musculoskeletal and Dermatological Sciences, Faculty of Medicine, Biology and
Health, University of Manchester, Manchester, UK, ³³²Department of Biostatistics, University of
Liverpool, Liverpool, UK, ³³³MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

* Denote shared authorship contributions

Abstract

Glycaemic traits are used to diagnose and monitor type 2 diabetes, and cardiometabolic health. To date, most genetic studies of glycaemic traits have focused on individuals of European ancestry. Here, we aggregated genome-wide association studies in up to 281,416 individuals without diabetes (30% non-European ancestry) with fasting glucose, 2h-glucose post-challenge, glycated haemoglobin, and fasting insulin data. Trans-ancestry and single-ancestry meta-analyses identified 242 loci (99 novel; $P < 5 \times 10^{-8}$), 80% with no significant evidence of between-ancestry heterogeneity. Analyses restricted to European ancestry individuals with equivalent sample size would have led to 24 fewer new loci. Compared to single-ancestry, equivalent sized trans-ancestry fine-mapping reduced the number of estimated variants in 99% credible sets by a median of 37.5%. Genomic feature, gene-expression and gene-set analyses revealed distinct biological signatures for each trait, highlighting different underlying biological pathways. Our results increase understanding of diabetes pathophysiology by use of trans-ancestry studies for improved power and resolution.

Introduction

Fasting glucose (FG), 2h-glucose post-challenge (2hGlu), and glycated haemoglobin (HbA1c) are glycaemic traits used to diagnose diabetes¹. In addition, HbA1c is the most commonly used biomarker to monitor glucose control in patients with diabetes. Fasting insulin (FI) reflects a combination of insulin resistance, a component of type 2 diabetes (T2D), and insulin clearance². Collectively, all four of these glycaemic traits can be useful to better understand T2D pathophysiology³⁻⁵, are useful measures of cardiometabolic health as they are associated with cardiometabolic outcomes even within the non-diabetic range, albeit modestly so⁶.

To date, genome-wide association studies (GWAS) and analysis of next-generation targeted arrays (MetaboChip and exome array) have identified >120 loci associated with glycaemic traits in individuals without diabetes⁷⁻¹⁵. However, despite considerable differences in the prevalence of T2D risk factors across ancestries¹⁶⁻¹⁸, most glycaemic trait GWAS in individuals without diabetes have insufficient representation of individuals of non-European ancestry and limited resolution for fine-mapping of causal variants and effector transcript identification. Here, we present large-scale trans-ancestry discovery meta-analyses of GWAS for four glycaemic traits (FG, 2hGlu, FI, and HbA1c) in individuals without diabetes with genotype imputation to the 1000 Genomes Project reference panel phase 1 version 3¹⁹. Our aims were to identify additional glycaemic trait-associated loci; investigate the portability of loci and genetic scores across ancestries; leverage differences in effect allele frequency (EAF), effect size, and linkage disequilibrium (LD) across diverse populations to conduct fine-mapping and aid causal variant/effector transcript identification; and compare and contrast the genetic architecture of these four glycaemic traits to further elucidate their underlying biology and gain insights into pathophysiological pathways implicated in T2D.

Results

Study design, lead variant, index variant and trans-ancestry locus definitions

To identify loci associated with glycaemic traits FG, 2hGlu, FI, and HbA1c, we aggregated GWAS in up to 281,416 individuals without diabetes, ~30% of whom were of non-European ancestry [13% East Asian, 7% Hispanic, 6% African-American, 3% South Asian, and 2% sub-Saharan African (Ugandan - data only available for HbA1c)]. Prior to meta-analysis each contributing cohort imputed data to the 1000 Genomes Project reference panel (phase 1 v3, March 2012, or later; **Methods, Supplementary Table 1, Supplementary Figure 1**). In total, up to ~49.3 million variants were directly genotyped or imputed, with between 38.6 million (2hGlu) and 43.5 million variants (HbA1c) available for analysis after exclusions based on minor allele count (MAC < 3) and imputation quality (imputation r^2 or INFO score < 0.40) in each cohort. As we had previously found adjusting for body mass index (BMI) provided similar results for FG and 2hGlu, but aided in new locus discovery for FI¹⁵, here we conducted analyses for FG, 2hGlu and FI adjusted for BMI, but for simplicity these traits are abbreviated as FG, 2hGlu and FI (**Methods**).

We first performed trait-specific fixed-effect meta-analyses *within* each ancestry using METAL²⁰. We defined “single-ancestry lead” variants as the strongest trait-associated variants ($P < 5 \times 10^{-8}$) within a 1Mb region in a particular ancestry (**Glossary box**). Within each ancestry and each autosome, we used approximate conditional analyses in GCTA^{21,22}, to identify distinct “single-ancestry index variants” ($P < 5 \times 10^{-8}$) that exert conditionally distinct effects on the trait (**Glossary Box, Methods**,

518 **Supplementary Figure 2).** Overall, this approach identified 124 distinct FG, 15 2hGlu, 48 FI and 139
 519 HbA1c variants that were significant in at least one ancestry (**Supplementary Table 2**).

Glossary Box

This study combined analyses of trait-associations across multiple correlated glycaemic traits and across multiple ancestries, which has presented challenges in our ability to apply commonly used terms with clarity. For this reason, we define below terms often used in the field with variable meaning, as well as definitions of new terms used in this study.

EA – the effect allele was that defined by METAL based on trans-ancestry FG results and aligned such that the same allele was kept as the effect allele across all ancestries and traits, irrespective of its allele frequency or effect size for that particular ancestry and trait, in this way the effect allele is not necessarily the trait-increasing allele.

Single-ancestry lead variant – variant with the smallest p-value amongst all with $P < 5 \times 10^{-8}$, within a 1Mb region, based on analysis of a single trait in a single ancestry.

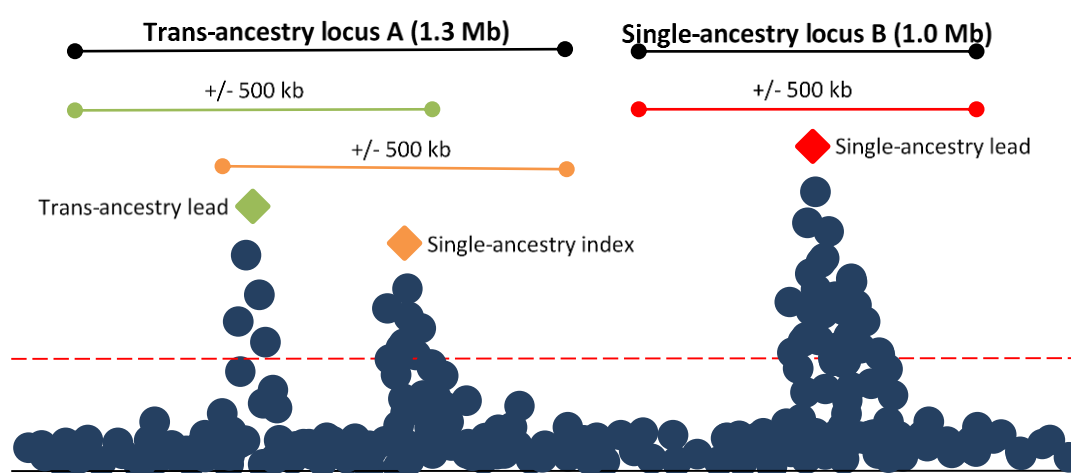
Single-ancestry index variants – variants identified by GCTA analysis of each autosome, and that appear to exert conditionally distinct effects on a given trait in a given ancestry ($P < 5 \times 10^{-8}$). As defined, these include the single-ancestry lead variant.

Trans-ancestry lead variant – variant identified by trans-ethnic meta-analysis of a given trait that has the strongest association for that trait ($\log_{10}BF > 6$, which is broadly equivalent to $P < 5 \times 10^{-8}$) within a 1Mb region.

Single-ancestry locus – a 1Mb region centred on a single-ancestry lead variant which does not contain a lead variant identified in the trans-ancestry meta-analysis (i.e., does not contain a trans-ancestry lead variant).

Signal – a conditionally independent association between a trait and a set of variants in LD with each other and which is noted by the corresponding index variant.

Trans-ancestry locus – As we expected some genetic variants to influence multiple correlated traits and that functional variants would influence traits across multiple ancestries, we combined results across traits and across ancestries into multi-trait trans-ancestry loci. A **trans-ancestry locus** is a genomic interval that contains trans-ancestry trait-specific lead variants, with/out additional single-ancestry index variants, for one or more trait. This region is defined by starting at the telomere of each chromosome and selecting the first single-ancestry index variant or trans-ancestry lead variant for any trait. If other trans-ancestry lead variants or single-ancestry index variants mapped within 500kb of the first signal, then they were merged into the same locus. This process was repeated until there were no more signals within 500kb of the previous variant. A 500kb interval was added to the beginning of the first signal, and the end of the last signal to establish the final boundary of the trans-ancestry locus. As defined, a trans-ancestry locus may not have a single lead trans-ancestry variant, but may instead contain multiple trans-ancestry lead variants, one for each trait.



Locus diagram– In this diagram, trans-ancestry locus A contains a trans-ancestry lead variant for one glycaemic trait represented by the green diamond, and another single-ancestry index variant for another glycaemic trait represented by the orange triangle. Single-ancestry locus B contains a single-ancestry lead variant represented by the red square. The orange, green and red bars represent a +/- 500Kb window around the orange, green, and red variants, respectively. The black bars indicate the full locus window where trans-ancestry locus A contains trans-ancestry lead and single-ancestry index variants for two traits and single-ancestry locus B has a single-ancestry lead variant for a single trait.

Next, we conducted trait-specific *trans-ancestry* meta-analyses of ancestry-specific results using MANTRA (**Methods, Supplementary Table 1, Supplementary Figures 1 and 3**) to identify genome-wide significant “trans-ancestry lead variants”, defined as the most significant trait-associated variant across all ancestries (\log_{10} Bayes Factor [BF] >6, equivalent to $P < 5 \times 10^{-8.23}$) (**Glossary box, Methods**). Here, we present trans-ancestry results based on data from all participating cohorts as our primary results (**Supplementary Table 2**).

Causal variants are expected to affect multiple related glycaemic traits and may be shared across ancestries. Therefore, we combined all single-ancestry lead variants, single-ancestry index variants, and/or trans-ancestry lead variants (for any trait) mapping within 500Kb of each other, into a single “trans-ancestry locus” that was bounded by a 500Kb flanking sequence (**Glossary Box**). As defined, a trans-ancestry locus may contain multiple causal variants affecting one or more glycaemic traits, exerting their effect in one or more ancestry.

Glycaemic trait locus discovery

In the trans-ancestry meta-analyses, we observed genome-wide significant associations at 235 trans-ancestry loci, of which 59 contained trans-ancestry lead variants for more than one trait. In addition, we identified seven “single-ancestry loci” that did not contain any trans-ancestry lead variants (**Glossary box, Supplementary Table 2**). Of the 242 trans-ancestry and single-ancestry loci, 99 (including 6 of the 7 single-ancestry) had not been previously associated with any of the four glycaemic traits or with T2D, at the time of analysis (**Figure 1, Supplementary Figures 1 and 3, Supplementary Table 3, Supplementary note**). Based on the largest European and East Asia ancestry T2D GWAS meta-analyses^{23,24}, the lead variants at 19 novel glycaemic trait loci have strong evidence of association with T2D ($P < 10^{-4}$; six loci with $P < 5 \times 10^{-8}$), suggesting some of the novel loci are also important in diabetes pathophysiology (**Supplementary Tables 2 and 4**).

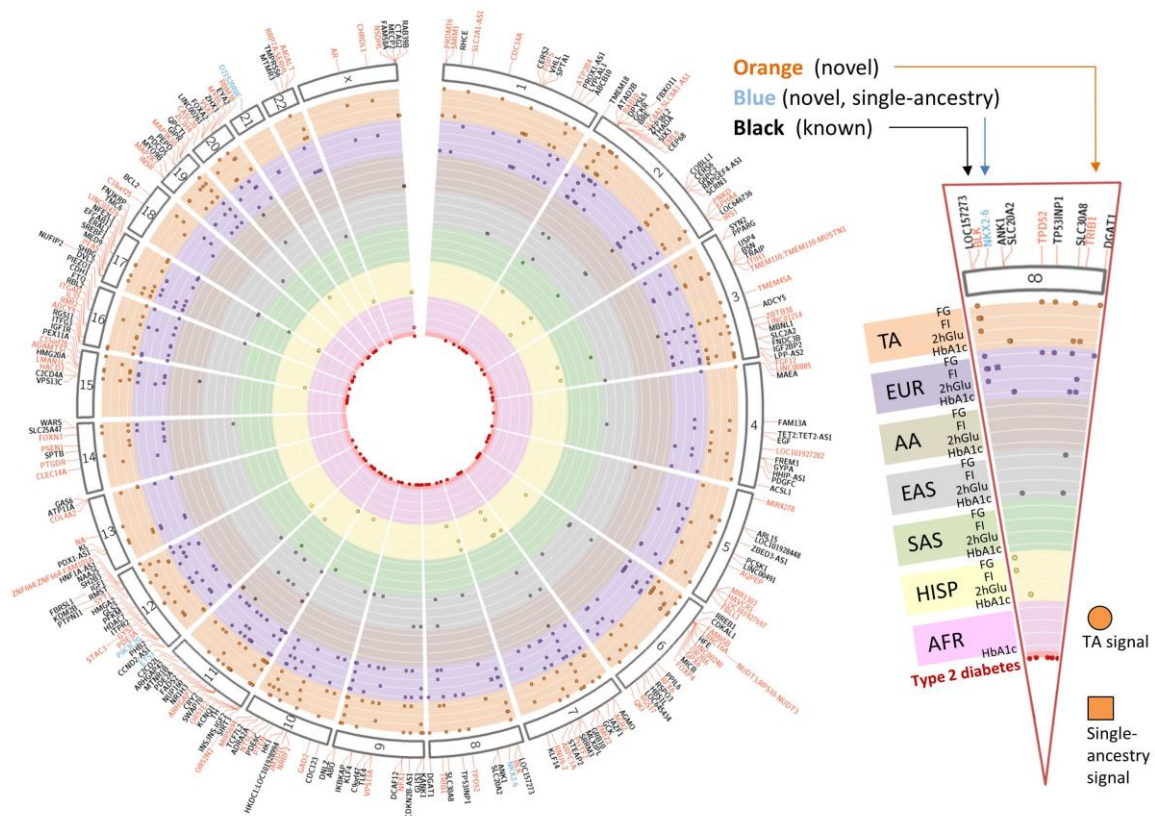


Figure 1 - Summary of all 242 loci identified in this study. 235 trans-ancestry loci are shown in orange (novel) or black (established) along with seven single-ancestry loci (blue) represented by nearest gene. Each locus is mapped to corresponding chromosome (outer segment). Each set of rows shows the results from the trans-ancestry analysis (orange) and each of the ancestries: European (purple), African American (tan), East Asian (grey), South Asian (green), Hispanic (yellow), sub-Saharan African (Ugandan-pink). Loci with a corresponding type 2 diabetes signal are represented by red circles in the middle of the plot.

Of the 99 novel loci, six were identified in a single ancestry and did not overlap a trans-ancestry locus (**Supplementary Table 3**). Three single-ancestry loci were associated in individuals of non-European ancestry: (i) an African American association for FG (lead variant rs61909476) near the gene *ETS1*, (ii) an African American association for FI (lead variant rs12056334) near the gene *LOC100128993* (an uncharacterised RNA gene; **Supplementary Note**), and (iii) a Hispanic association for FG (lead variant rs12315677) within the gene *PIK3C2G* (**Supplementary Table 3**). The associations of rs61909476 and rs12315677 with FG are noteworthy. The variant rs61909476 has a similar EAF in both European (~10%) and African American (~7%) ancestry populations (**Supplementary Table 2**), but the effect on FG is only detectable in African American individuals ($\beta=0.0812$ mmol/l, $SE=0.01$ mmol/l, $P=3.9 \times 10^{-8}$, European individuals $\beta=0.0015$ mmol/l, $SE=0.0031$ mmol/l, $P=0.44$) (**Supplementary Figure 4, Supplementary note**). The nearest gene, *ETS1*, encodes a transcription factor which has been shown to localize to insulin-positive cells in mouse islets, and its overexpression was shown to decrease glucose-stimulated insulin secretion in mouse islets²⁵. Located within the *PIK3C2G* gene, rs12315677 has a similar EAF in both Hispanic and European ancestry populations (84% and 86%, respectively), but is significantly associated with FG only in our

Hispanic GWAS ($\beta=0.0387$ mmol/l, $SE=0.0075$ mmol/l, $P=4.0\times 10^{-8}$) compared with European ancestry ($\beta=-0.0029$ mmol/l, $SE=0.0029$ mmol/l, $P=0.39$) (**Supplementary Figure 5, Supplementary note**). *PIK3C2G* has been shown to be a Rab5 effector which, when deleted in *Pik3c2g*^{-/-} mice, selectively inhibits *Akt2* activation and leads to a phenotype characterised by reduced glycogen storage in the liver, hyperlipidaemia, adiposity, and insulin resistance with increasing age, or after a high fat diet²⁶. Instances where the EAFs are similar between populations, but the effect sizes differ, could be due to specific genotype-by-environment effects that differ across ancestries, or lower imputation accuracy in ancestries with smaller sample sizes, although this would likely lead to deflated effect sizes and imputation quality is good for these variants (average $r^2=0.81$). It is also possible that the variants detected here are not themselves causal, but are in LD with ancestry-specific causal variants that are not directly interrogated in our meta-analysis and that differ in frequency across ancestries. To try and investigate this hypothesis, we looked at data from 1000G in the cognate populations for evidence of rarer alleles in those ancestries that may themselves be driving the association signals (**Supplementary Table 5**). We could not detect evidence for other rarer alleles driving these associations, but this does not preclude the possibility that other rarer variants exist which are not represented in the 1000G populations. The final three single-ancestry loci were identified in individuals of European ancestry, but without any evidence of association in the other ancestries despite similar MAF, although this may be due to differences in power given the much smaller sample sizes in non-European ancestries (**Supplementary Figures 6-8**).

Next, we investigated the contribution of non-European ancestry data to novel trans-ancestry locus discovery, independent of the total sample size in the trans-ancestry meta-analysis. To do this, we artificially boosted the sample size of the European meta-analysis to match that of trans-ancestry meta-analysis by rescaling the standard errors of allelic effect sizes (**Supplementary note**). Using this approach, we determined that 21 of the novel trans-ancestry loci would not have been discovered if the sample size obtained in the trans-ancestry analyses was comprised exclusively of European ancestry individuals (**Supplementary note**). Instead, their discovery was due to the higher EAF and/or larger effect size in non-European ancestry populations. In particular, two loci (nearest genes *LINC00885* and *MIR4278*) contain East Asian and African American single-ancestry lead variants, respectively, suggesting that these specific ancestries may be driving the trans-ancestry discovery (**Supplementary Tables 2-3**). Combined with the three single-ancestry non-European loci described above, our results show that 24% (24/99) of novel loci were discovered due to the contribution of non-European ancestry participants, strengthening the argument for extending genetic studies to larger samples sizes in diverse populations.

Allelic architecture of glycaemic traits

Trans-ancestry and single-ancestry loci comprised a range of association patterns, with most loci harbouring one single-ancestry signal for any given trait (**Supplementary note**). However, 29 loci contained multiple distinct index variants that did not fully overlap between ancestries. The most complex locus we observed was in the region spanning *G6PC2*, which contained 14 distinct FG index variants in the European single-ancestry meta-analysis. Of these, four are shared ($P<5\times 10^{-8}$) with South Asian ancestry, two with East Asian ancestry, and two with Hispanic ancestry (**Supplementary Figure 9**). The complexity of association signals at this locus is consistent with previous work that

also reported common variant (MAF>5%) association signals and multiple rare variant (MAF≤1%) associations at this locus that influenced protein function by multiple mechanisms²⁷.

Combined, single-ancestry lead, single-ancestry index, and trans-ancestry lead variants increase the number of established loci for FG to 102 (182 signals, 53 novel loci), FI to 66 (95 signals, 49 novel loci), 2hGlu to 21 (28 signals, 11 novel loci), and HbA1c to 127 (218 signals, 62 novel loci) (**Supplementary Table 2**) and demonstrate significant overlap across glycaemic traits (**Supplementary Figure 10**). We also detected ($P<0.05$ or $\log_{10}BF>0$) the vast majority (~90%) of previously established glycaemic trait association signals in our data, 70-88% of which attained genome-wide significance in the current analyses (see further details in the **Supplementary Note**). Given that analyses for FG, FI, and 2hGlu were performed adjusted for BMI, we also confirmed that collider bias was not influencing discovery for more than 98% of our results (**Supplementary note**)²⁸.

Finally, as expected, given the greater power due to increased sample sizes, new association signals tended to have smaller effect sizes and/or EAFs in European ancestry individuals (in whom this analysis was conducted) compared to previously established signals (**Supplementary Figure 11**).

Characterisation of trans-ancestry lead variants and European index variants across ancestries

We next employed a series of complementary analyses to better understand the transferability of trans-ancestry lead variants across all ancestries. For each trans-ancestry lead variant, we investigated the pairwise EAF correlation between ancestries, as well as the pairwise summarised heterogeneity of effect sizes between ancestries²⁹ (**Methods** and **Supplementary Note**). In agreement with population history and evolution, these results demonstrated considerable EAF correlation ($\rho^2>0.70$) between European and Hispanic populations, European and South Asian populations, and Hispanic and South Asian populations, consistent across all four traits, and between African Americans and Ugandans for HbA1c (**Supplementary Figure 12**). Despite significant EAF correlations, some pairwise comparisons exhibited strong evidence for effect size heterogeneity between ancestries that was less consistent between traits (**Supplementary Figure 12**). However, sensitivity analyses demonstrated that, across all comparisons, the evidence for heterogeneity is driven by a small number of variants, with between 81.5% (for HbA1c) and 85.7% of trans-ancestry lead variants (for FG) showing no evidence for trans-ancestry heterogeneity ($P>0.05$) (**Supplementary Note**).

We also took LD pruned European single-ancestry index variants and compared the direction of effect of these variants in European ancestry individuals with that in other ancestries (**Supplementary Note**). Consistent with the lack of heterogeneity in effect sizes, we saw >70% concordance in the direction of effect for all traits into all ancestries, with the exception of HbA1c into African Americans and Ugandans (**Supplementary Table 6**). Imperfect concordance between ancestries could reflect lower power in non-European ancestry groups due to sample size or variation in allele frequency, or could be explained by LD differences between index SNPs and causal variants. For HbA1c, we hypothesized that lower concordance might also be a reflection of the different pathways (glycaemic and non-glycaemic) through which variants can affect HbA1c levels,

particularly effects mediated via the red blood cell (RBC) where balancing selection can lead to different associations in individuals of African ancestry⁷ (**Supplementary Note** and below).

To further investigate the potential utility of trans-ancestry analyses, and to evaluate whether larger sample sizes might yield additional European ancestry signals that would be transferable across ancestries, we extended these concordance analyses to the entire genome, clumping variants mapping >1Mb apart (to eradicate the effect of LD in all ancestries) in different bins of association p-values obtained from the European ancestry meta-analysis (**Methods**). Aside from the bins with the weakest evidence for association in Europeans (i.e. in all bins with $P \leq 0.05$), we observed nominally significant concordance in the direction of effects between European and other ancestries for all traits except for 2hGlu, in which analyses were underpowered (**Supplementary Table 6**).

Transferability of genetic scores (GS) across ancestries

To investigate the portability of GS across ancestries (the equivalent of genetic risk scores used for disease studies but instead for quantitative traits), we built a GS on the basis of effect sizes at European single-ancestry index variants ($P < 5 \times 10^{-8}$), after LD pruning ($r^2 < 0.1$), and assessed its utility for predicting trait variance explained in other ancestries (**Methods, Supplementary Note, Supplementary Table 7**). As a benchmark, we first assessed the predictive power (trait variance explained, as assessed by R^2) of the GS into each cohort contributing to the European meta-analysis and three additional European cohorts that were not part of the meta-analysis. We then assessed the trait variance explained by the GS into the other ancestries and observed that the R^2 fell within the range of values observed across European cohorts (**Figure 2A**).

We next expanded the GS to include all distance-based clumped variants across the genome with nominal evidence of association ($P < 1 \times 10^{-5}$) in European ancestry individuals. This expansion improved the trait variance explained (greater R^2) of the GS into European ancestry individuals compared to the GS built from LD pruned single-ancestry European lead and index variants ($P < 1 \times 10^{-8}$) but substantially worsened performance into other ancestries (**Supplementary Table 8**).

Finally, using GS with a combination of individual trait trans-ancestry lead variants and single-ancestry lead and index variants within each ancestry, we were able to demonstrate that these explained, on average, between 0.7% (2hGlu in EUR) and 6% (HbA1c in AA) of the variance in trait distribution (**Methods, Figure 2B, Supplementary Tables 9-12**). In Europeans, these estimates represent an improvement (i.e. more variance explained) relative to previous estimates, derived using similar methodology, of 2.8% for FG and 1.7% for HbA1c³⁰. Whilst variance explained estimates of 4.8% (FG), 1.2% (FI) and 1.7% (2hGlu) reported by Scott et al¹⁵ are in excess of our estimates, we hypothesise this is likely to be at least partly attributable to a difference in statistical approaches (see further discussion in **Supplementary Note**).

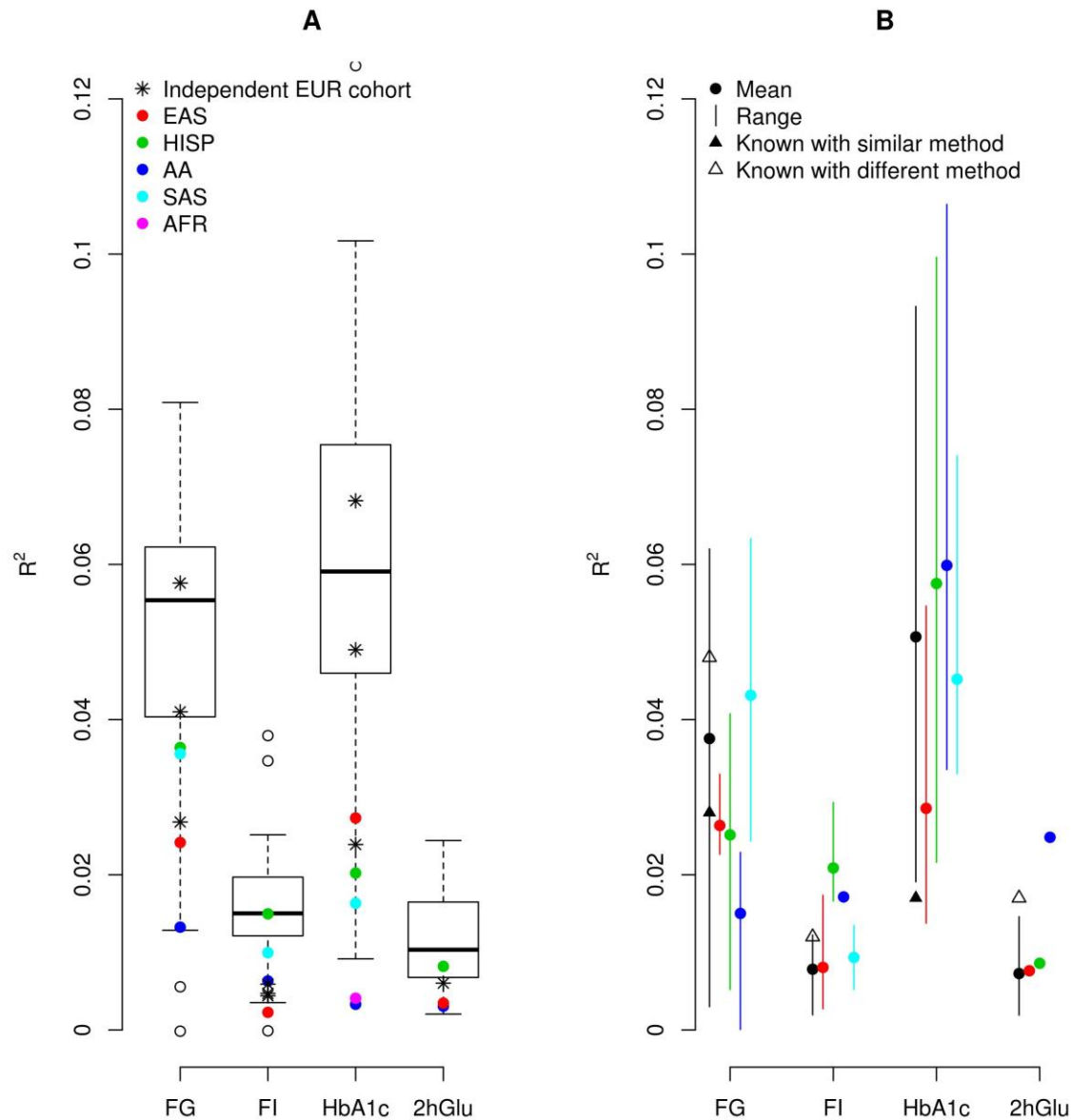
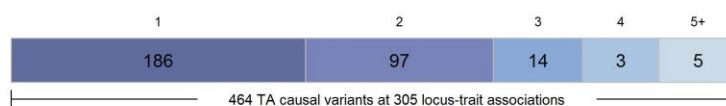


Figure 2 – Transferability of GS across ancestries. Coloured dots represent data from the different ancestries: EUR in black, EAS in red, HISP in green, AA in blue and SAS in light blue. A) represents trait variance explained (FG, FI, HbA1c and 2hGlu) in each ancestry based on a GS build on the basis of effect sizes at European single-ancestry lead and index variants, after LD pruning ($r^2 < 0.1$). The boxplot shows the maximum, first quantile, median, third quantile and minimum of variance explained in the EUR ancestry cohorts included in the study. The black asterisks show additional EUR cohorts that were not part of the original meta-analysis, while the dots represent variance explained in each of the other ancestries. B) represents trait variance explained when using a GS with a combination of individual trait trans-ancestry lead variants and single-ancestry lead and index variants, within each ancestry. Variance explained (mean and range of R^2) for each trait (FG, FI, HbA1c, and 2hGlu) in each ancestry is shown. R^2 was estimated in 1 to 11 cohorts with sample sizes ranging from 489 to 9,758 (**Supplementary Tables 9-12**). Closed and open triangles display previous known results using a similar method³⁰ or a different method¹⁵.

Fine-mapping

Of the 242 identified loci, 231 were autosomal trans-ancestry loci and six were autosomal single-ancestry loci, which we took forward for fine-mapping (**Supplementary Table 2**). Due to the absence of LD maps from adequately sized populations, fine-mapping was not attempted for the 5 loci (4 trans-ancestry and 1 single-ancestry) mapping to the X chromosome. Using FINEMAP with ancestry-specific LD and an average LD matrix across ancestries, we conducted fine-mapping both within single-ancestries (all 237 autosomal loci) and across ancestries (231 autosomal trans-ancestry loci) for each trait (**Methods**). Because 59 of the 231 trans-ancestry loci were associated with more than one trait, we conducted trans-ancestry fine-mapping for a total of 305 locus-trait associations. Of these 305 locus-trait combinations, FINEMAP estimated the presence of a single causal variant responsible for the association at 186 loci (61%), while multiple distinct causal variants were implicated at 126 loci (39%), for a total of 464 causal variants (**Figure 3A**).

A. Estimated number of causal variants at each locus-trait association



B. Number of variants at each 99% credible set



C. Fine-mapping resolution

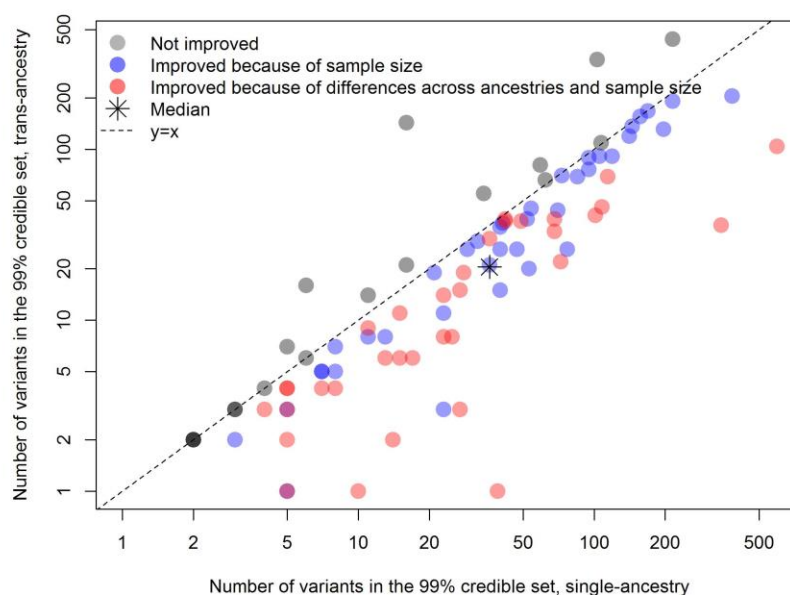


Figure 3 - Trans-ancestry fine-mapping. A) Number of plausible causal variants at each locus-trait association derived from FINEMAP. B) Number of variants within each 99% credible set. Twenty-one locus-trait associations at 19 loci were mapped to a single variant in the 99% credible set. C) Fine-mapping resolution. For each of the 98 locus-trait associations with a predicted single causal variant in both trans-ancestry and single-ancestry analyses, the number of variants included in the 99% credible set in the single-ancestry fine-mapping (x axis; logarithmic scale) is plotted against those in the trans-ancestry fine-mapping (y axis; logarithmic scale). Trans-ancestry and single-ancestry fine-mapping were based on the same set of variants. After removing eight locus-trait associations with one variant in the 99% credible sets in both trans-ancestry and single-ancestry analyses, there were 18 locus-trait associations (in grey) where trans-ancestry fine-mapping did not improve the resolution of fine-mapping results (i.e. number of variants in the 99% credible set did not decrease). Of the 72 locus-trait associations with improved trans-ancestry fine-mapping resolution (blue and red) further analyses in European fine-mapping emulating the total sample size in trans-ancestry fine-mapping demonstrated that 34 locus-trait associations (in red) were improved because of both total sample size and differences across ancestries, while 38 locus-trait associations (in blue) were only improved due to increased sample size in the original trans-ancestry fine-mapping analysis.

Credible sets for causal variants

At each locus, we next constructed credible sets (CS) for each causal variant that account for $\geq 99\%$ of the posterior probability of association (PPA). We identified 21 locus-trait associations (at 19 loci) for which the 99% CS included a single variant, and we highlight five examples below. (**Methods, Supplementary Note, Figure 3B, Supplementary Table 13**).

First, we highlight two positive controls which provide confidence in the results. At one locus near *MTNR1B*, rs10830963 (PPA >0.999 , for both HbA1c and FG), located in an *MTNR1B* intron, has shown allelic differences in enhancer activity and transcription factor binding³¹. At an additional FG-associated locus near *SIX3*, rs12712928 (PPA=0.997) has shown allelic differences in transcriptional activity, transcription factor binding, and association with islet expression levels of nearby genes *SIX3* and *SIX2*^{32,33}. The EAF and effect size of this variant is larger in EAS than in other ancestries (heterogeneity p-value=7.2x10⁻⁸), which is driving the association at this locus.

Second, we highlight three novel findings. At a locus near *PFKM* associated with HbA1c, trans-ancestry fine-mapping identified rs12819124 (PPA >0.999) as the likely causal variant. This variant has been previously associated with mean corpuscular haemoglobin³⁴, suggesting an effect of this locus on HbA1c is via the RBC. We note that this locus also harbours an association with FI in European and trans-ancestry meta-analyses, although it appears to be distinct from the HbA1c signal based on distance and LD. Fine-mapping of the nearby FI signal in European ancestry populations identified rs111264094 (PPA=0.994) as the likely causal variant (**Supplementary Figures 13-14**). rs111264094 is a low frequency variant in Europeans (EAF=0.025) that is monomorphic or rare in other ancestries, is located >600 kb from HbA1c-associated variant rs12819124, and is in low LD with rs12819124 in European ancestry populations ($r^2<0.1$), which supports the hypothesis of two distinct signals (one for FI and one HbA1c) at this locus.

At the *HBB* locus, we also identify rs334 (PPA >0.999 ; Glu7Val) as the likely causal variant associated with HbA1c. rs334 is a causal variant of sickle cell anaemia³⁵, with previously reported associations with urinary albumin-to-creatinine ratio in Caribbean Hispanic individuals³⁶, severe malaria in a

Tanzanian study population³⁷, haematocrit and mean corpuscular volume in Hispanic/Latino populations³⁸, and more recently with RBC distribution in Ugandan individuals³⁹, all of which point to an effect of this variant on HbA1c via non-glycaemic pathways.

Lastly, our credible set analysis identified rs1799815 (PPA=0.993) as the likely causal variant at the *INSR* locus associated with FI. rs1799815 is a synonymous variant (Tyr3033Tyr) within *INSR*, the well-known insulin receptor gene that regulates the insulin signalling pathway. *INSR* as a target gene for this locus is further supported by our finding that rs1799815 colocalizes as an eQTL for *INSR* expression in adipose tissue (details shown below). The remaining locus-trait associations with a single variant in the 99% CS (**Supplementary Table 13**) point to variants that could be prioritised for downstream functional follow-up to further elucidate their impact on glycaemic trait physiology.

In addition to identifying 99% CS with a single variant, trans-ancestry fine-mapping identified 99% CS with 50 or fewer variants at 156 locus-trait associations (**Figure 3B, Supplementary Table 13**). Overall, 74 locus-trait associations contained 87 variants with PPA>0.90; that is, some locus-trait associations contain more than one variant with a high predicted probability of being causal as there can be more than one causal variant in a locus (**Supplementary Table 14**). In addition to those already described above, the identified variants are strong candidate causal variants that merit prioritisation for future functional validation. For example, among the 87 variants, 10 are coding variants including several missense such as the *HBB* Glu7Val mentioned above, *GCKR* Leu446Pro, *RREB1* Asp1771Asn, *G6PC2* Pro324Ser, *GLP1R* Ala316Thr, and *TMPRSS6* Val736Ala, each of which have been proposed or shown to affect gene function^{12,40-44}. We also additionally identify *AMPD3* Val311Leu (PPA=0.989) and *TMC6* Trp125Arg (PPA>0.999) variants associated with HbA1c which were previously detected in an exome array analysis but had not been fine-mapped with certainty due to the absence of backbone GWAS data²⁷. Our current fine-mapping data now suggest these variants are likely to be causal and identify the cognate genes as the effector transcripts driving these associations.

Finally, we evaluated the resolution obtained in the trans-ancestry versus single-ancestry fine-mapping (**Methods, Supplementary Note**). To do this, we compared the number of variants in 99% CS across 98 locus-trait associations which, as suggested by FINEMAP, had a single causal variant in both trans-ancestry and single-ancestry analyses. Fine-mapping within and across ancestries was conducted using the same set of variants. At 8 of 98 locus-trait associations single-ancestry fine-mapping identified a single variant in the CS. In addition, at 72 of the 98 locus-trait associations, the number of variants in the 99% CS was smaller in trans-ancestry fine-mapping than in single-ancestry analyses (**Figure 3C**), which likely reflects the larger sample size and differences in LD structure, EAFs, and effect sizes across diverse populations. To quantify the estimated improvement in fine-mapping resolution attributable to the multi-ancestry GWAS, we then compared 99% CS sizes from the trans-ancestry fine-mapping to single-ancestry-specific data emulating the same total sample size by rescaling the standard errors (**Methods**). Of the 72 locus-trait associations with estimated improved fine-mapping in trans-ancestry analysis, resolution at 38 (53%) was improved because of the larger sample size in the trans-ancestry fine-mapping analysis (**Figure 3C**), and this estimated improved resolution would likely have been obtained in a European-only fine-mapping effort with equivalent sample size. However, at 34 (47%) loci, the inclusion of samples from multiple diverse

populations yielded estimated improved resolution. On average, ancestry differences led to a reduction in the median number of variants in the 99% CS from 24 to 15 variants (37.5% median reduction; **Figure 3C**), demonstrating the value of conducting fine-mapping across ancestries.

HbA1c Signal Classification

We, and others, have previously suggested that HbA1c-associated variants appear to exert their effects on HbA1c levels through both glycaemic and non-glycaemic pathways^{7,45}. Classification of loci into these pathways can have important implications for T2D diagnostic accuracy^{7,46}. To further elucidate the biology of HbA1c-associated variants, we took advantage of prior association results for other glycaemic, RBC, and iron traits, and used a fuzzy clustering approach to classify variants into their most likely mode of action (**Methods, Supplementary note**). Of the 202 autosomal HbA1c-associated trans-ancestry lead variants and single-ancestry index variants, 16 (8%) could not be characterized due to missing summary statistics in the other datasets and 17 (8%) could not be classified into a “known” class (**Supplementary note**). The remaining signals were classified as principally: a) glycaemic (n=51; 25%), b) affecting iron levels/metabolism (n=12; 6%), or c) RBC traits (n=106; 53%). We found a genetic risk score (GRS) composed of all HbA1c-associated signals was strongly associated with T2D risk (OR=2.5, 95% CI 2.5-2.6, $P=2.4 \times 10^{-301}$). However, when we tested partitioned GRSs composed of these different classes of variants (**Methods**), we found the T2D association was mainly driven by those variants influencing HbA1c through glycaemic pathways (OR=2.8, 95% CI 2.7-2.9, $P=1.1 \times 10^{-251}$), with weaker evidence of association (despite the larger number of variants in the GRS) and a more modest risk (OR=1.4, 95% CI 1.3-1.5, $P=6.9 \times 10^{-4}$) imparted by signals in the mature RBC cluster that were not glycaemic (i.e. where those specific variants had $P>0.05$ for FI, 2hGlu and FG) (**Supplementary Figure 15, Supplementary note**). This contrasts our previous finding where we found no significant association between a risk score of non-glycaemic variants and T2D⁷. Our current results could be partly driven by T2D cases being diagnosed based on HbA1c levels that may be influenced by the non-glycaemic signals, or by glycaemic effects not captured by FI, 2hGlu or FG measures.

Biological signatures of glycaemic trait associated loci

To better understand distinct and shared biological signatures underlying variant-trait associations, we conducted genomic feature enrichment, eQTL co-localisation, and tissue and gene-set enrichment analyses across all four traits.

Epigenomic landscape of trait-associated variants

We next explored the genomic context underlying glycaemic trait loci by computing overlap enrichment for static annotations such as coding, conserved regions, histone modification ChIP-seq peaks, and super enhancers, merged across various cell types⁴⁷⁻⁴⁹ using the GREGOR tool⁵⁰. We observed that FG, FI and HbA1c signals (**Supplementary Table 7**) were significantly ($P<8.4 \times 10^{-4}$, Bonferroni threshold correcting for 59 total annotations) enriched in evolutionarily conserved regions, whereas 2hGlu signals were only nominally enriched (**Fig 4A, Supplementary Figure 16, Supplementary Table 15**).

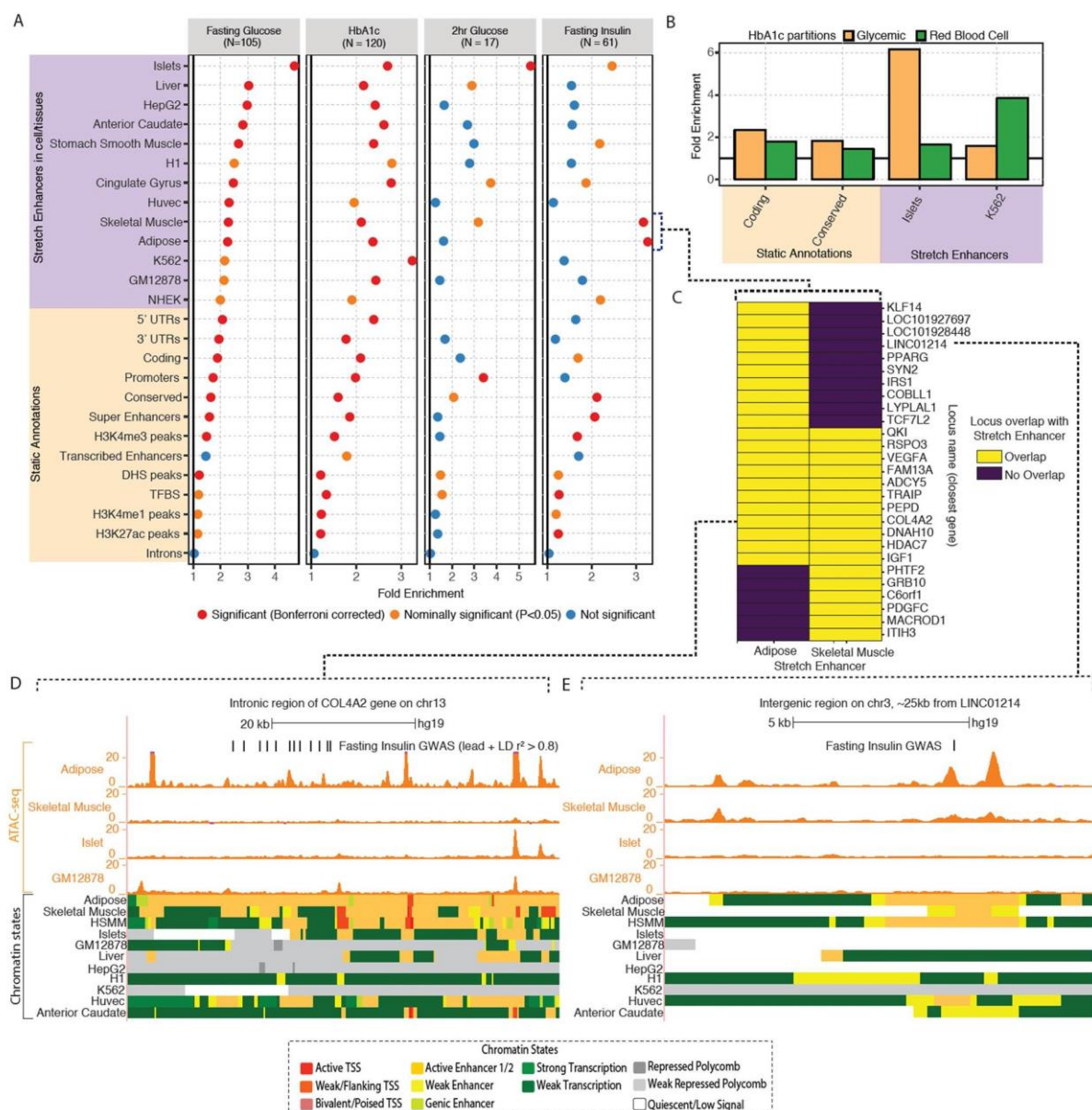


Figure 4 - Epigenomic landscape of trait-associated variants. A: Enrichment of GWAS variants to overlap genomic regions including 'Static Annotations' which are common or 'static' across cell types and 'Stretch Enhancers' which are identified in each tissue/cell type. The numbers of signals for each trait are indicated in parentheses. Enrichment was calculated using GREGOR⁵⁰. Significance (red) is determined after Bonferroni correction to account for 59 total annotations tested for each trait; nominal significance ($P < 0.05$) is indicated in yellow. B: Enrichment for HbA1c GWAS signals partitioned into "hard" Glycaemic and Red Blood Cell cluster (signals from "hard" mature Red Blood Cell and reticulocyte clusters together) to overlap annotations including stretch enhancers in Islets and the blood-derived leukemia cell line K562, respectively (additional partitioned results in **Supplementary Table 17**). C: Individual FI GWAS signals that drive enrichment in Adipose and Skeletal Muscle stretch enhancers. D, E: Genome browser shots of FI GWAS signals – intronic region of the *COL4A2* gene (D) and an inter-genic region ~25kb from *LINC01214* gene (E) showing GWAS SNPs (lead and LD $r^2 > 0.8$ proxies), ATAC-seq signal tracks and chromatin state annotations in different tissues/cell types.

We then focussed on the epigenomic landscapes defined in individual cell/tissue types. Previously, stretch enhancers (enhancer chromatin states $\geq 3\text{kb}$ in length) in pancreatic islets were shown to be highly cell-specific and strongly enriched with T2D risk signals⁵¹. We therefore calculated the enrichment of glycaemic trait-associated signals (**Supplementary Table 7**) in previously defined stretch enhancers³³ across a diverse panel of cell types and tissues most relevant to the traits of interest: pancreatic islets, skeletal muscle, adipose, and liver (**Methods**). These analyses strongly suggest that variants associated with these glycaemic traits influence the function of tissue specific enhancers. Namely, FG- and 2hGlu-associated signals have the highest enrichment in islet stretch enhancers (FG: fold enrichment=4.70, $P=2.7 \times 10^{-24}$; 2hGlu: fold enrichment=5.51, $P=3.6 \times 10^{-4}$ **Figure 4A, Supplementary Table 16**), which highlights the relevance of pancreatic islet tissue for the regulation of FG and 2hGlu. Interestingly, FI-associated variants are strongly enriched for overlap with stretch enhancers in skeletal muscle (fold enrichment=3.17, $P=7.8 \times 10^{-6}$) and adipose tissue (fold enrichment=3.27, $P=1.8 \times 10^{-7}$), which is consistent with these tissues being key targets of insulin action and their involvement in the insulin resistance phenotype (**Figure 4A**). We note that the high enrichment of stretch enhancers in individual cell types (see upper “stretch enhancer” labelled portion of **Figure 4A**) as compared to super enhancers merged across cell types (see lower “static annotations” labelled portion of **Figure 4A**) highlights the importance of using cell-specific annotations in enrichment analyses. HbA1c-associated signals are enriched in stretch enhancers of multiple cell types and tissues likely because of the complex nature of this trait, but have the strongest enrichment in stretch enhancers from the blood-derived leukaemia cell line K562 (fold enrichment=3.24, $P=1.21 \times 10^{-7}$, **Figure 4A**). We next sought to identify potential cell specific epigenomic enrichments that are associated with the classified HbA1c-associated variants corresponding to the “hard” glycaemic and red blood cell clusters, the latter being the joint group of mature red blood cell and reticulocyte clusters. We found that these partitioned variants display expected cell type-specific enrichment trends with the HbA1c glycaemic variants significantly enriched in islet stretch enhancers (fold enrichment=6.25, $P=4.02 \times 10^{-10}$ **Figure 4B, Supplementary Table 17**) and not in K562. Conversely, the HbA1c red blood cell variants are significantly enriched in K562 stretch enhancers (fold enrichment=3.85, $P=3.32 \times 10^{-8}$, **Figure 4B, Supplementary Table 17**) and not in islets.

To complement the overlap enrichment results from GREGOR, we also computed enrichment with two additional approaches: fGWAS⁵² and GARFIELD⁵³. These independent analyses yielded consistent results (**Supplementary Figures 17-18, Supplementary Tables 15 and 18**), demonstrating reproducibility across different approaches.

Given the observed enrichment of FI loci with stretch enhancers from adipose and skeletal muscle tissue, we sought to explore these loci in more detail. We found that 11 of the 27 loci driving these enrichment signals include variants that overlap stretch enhancers in both adipose and skeletal muscle (**Figure 4C**). At the *COL4A2* locus, variants within an intronic region of the gene overlap stretch enhancer chromatin states in adipose tissue, skeletal muscle, and a human skeletal muscle myoblast (HSMM) cell line that are not shared across other cell types and tissues; among these variants, rs9555695 (in the 99% CS) also overlaps accessible chromatin regions in adipose (**Figure 4D**). At a narrow signal (no proxy variants with LD $r^2 > 0.7$ in Europeans, for the lead trans-ancestry

rs62271373 variant), rs62271373 (PPA = 0.94) located in an intergenic region ~25kb from the *LINC01214* gene overlaps stretch enhancer chromatin states in adipose and HSMM and active enhancer chromatin states in skeletal muscle, but does not overlap any enhancer states in other tissues (**Figure 4E**). The lead rs62271373 variant also overlaps an ATAC-seq peak in adipose tissue. Collectively, the tissue-specific stretch enhancer epigenomic signatures at GWAS signals provide an opportunity to nominate tissues where these variants are likely to be active. Such a map will be helpful in future efforts to deconvolute GWAS signals into tissue-specific disease pathology.

Co-localisation of GWAS and eQTLs

Among the 99 novel glycaemic trait loci identified by this study, we identified co-localised eQTLs at 34 loci in blood, pancreatic islets, subcutaneous or visceral adipose, skeletal muscle, or liver, providing suggestive evidence of causal genes (**Supplementary Table 19**). The co-localised eQTLs include several genes previously reported at glycaemic trait loci: *ADCY5*, *CAMK1D*, *IRS1*, *JAZF1*, and *KLF14*⁵⁴⁻⁵⁶. For some additional loci, the co-localised genes have prior evidence for a role in glycaemic regulation. For example, the lead trans-ancestry variant and likely causal variant, rs1799815 (PPA=0.993, mentioned above), associated with FI is the strongest variant associated with expression of *INSR*, encoding the insulin receptor, in subcutaneous adipose from METSIM ($P=2 \times 10^{-9}$) and GTEx ($P=5 \times 10^{-6}$). The A allele at rs1799815 is associated with higher FI and lower expression of *INSR*, which is consistent with the well-established relationship in humans and model organisms between insulin resistance and reduced function of INSR protein⁵⁷. In a second example, rs841572, the trans-ancestry lead variant associated with FG, is the variant with the highest PPA (PPA=0.535) among the 20 variants in the 99% CS and is in strong LD ($r^2=0.87$) with the lead eQTL variant (rs841576, also in the 99% CS) associated with expression of *SLC2A1* in blood from eQTLGen ($P=1 \times 10^{-8}$). *SLC2A1*, also known as *GLUT1*, encodes the major glucose transporter in brain, placenta, and erythrocytes, and is responsible for glucose entry into the brain⁵⁸. The A allele at rs841572 is associated with lower FG and lower *SLC2A1* expression. While rare missense variants in *SLC2A1* are an established cause of seizures and epilepsy⁵⁹, our data suggest that *SLC2A1* variants also affect plasma glucose levels within a healthy physiological range. At both loci, the novel associations and co-localised eQTLs provide strong human genetic support for early glycaemia candidate genes.

The co-localised eQTLs also provide new insights into the mechanisms at glycaemic trait loci. For example, rs9884482 (a variant in the 99% CS) is associated with FI and expression of *TET2* in subcutaneous adipose ($P=2 \times 10^{-20}$); rs9884482 is in high LD ($r^2=0.96$ in Europeans) with the lead *TET2* eQTL variant (rs974801). *TET2* encodes a DNA-demethylase through which *TET2* can affect transcriptional repression⁶⁰. Adipose Tet2 expression is reduced in diet-induced insulin resistance in mice⁶¹, and knockdown of Tet2 blocked adipogenesis by repressing *Pparg* expression^{61,62}. Consistently, in human adipose tissue, rs9884482-C was associated with lower expression of *TET2* and higher FI. In a second example, HbA1c-associated variant rs617948 (a variant in the 99% CS) is the lead variant associated with expression of *C2CD2L* in blood from eQTLGen ($P=3 \times 10^{-96}$). *C2CD2L*, also known as *TMEM24*, has been shown to regulate pulsatile insulin secretion and facilitate release of insulin pool reserves^{63,64}. The G allele at rs617948 was associated with higher HbA1c and lower *C2CD2L*, providing evidence for a role of this insulin secretion protein in glucose homeostasis. Our HbA1c “soft” clustering classification assigns this signal to both the “unknown” (0.51 probability) and “reticulocyte” (0.42 probability) clusters, and this variant has no evidence for association with FG, FI or 2hGlu ($P>0.05$), but is strongly associated with HbA1c ($P<6.8 \times 10^{-8}$), reticulocytes (RET; $P<5 \times 10^{-7}$)

and HbA1c adjusted for FG ($P < 6.12 \times 10^{-7}$; **Supplementary Table 20, Supplementary Note**). Together, these results would suggest a possible effect of this variant on reticulocyte biology, and an effect on insulin secretion (mediated through *C2CD2L*) which is not captured by any of our traits, both of which potentially influencing HbA1c levels through different tissues, and providing a plausible explanation for the classification as “unknown”.

Tissue Expression

Consistent with results based on effector transcripts and expression analysis based on GTEx data ²⁷, we found significant differences in tissue expression across the glycaemic trait-associated variants. FG-associated variants were enriched for genes expressed in the pancreas (at FDR < 0.05), while there was insufficient power (insufficient number of genome-wide significant associations) in 2hGlu analysis to identify enrichment for any tissues or cell types at a more relaxed FDR < 0.2 threshold. FI-associated variants were enriched for connective tissue and cells (which includes adipose tissue), endocrine glands, blood cells, and muscles (at FDR < 0.2) and HbA1c-associated variants were significantly enriched for genes expressed in the pancreas, hemic, and immune system (at FDR < 0.05) (**Figure 5, Supplementary Table 21**). Consistent with our previous analysis ²⁷, FI-enrichment for connective tissue was driven by adipose tissue (subcutaneous and visceral), while the newly described enrichment with endocrine glands was driven by the adrenal glands and cortex (**Supplementary Table 21**). Beyond enrichment for genes expressed in glycaemic-related tissues, the association of HbA1c-associated variants with genes expressed in blood is consistent with the role of RBC in this glycaemic measure and our previous results ²⁷.

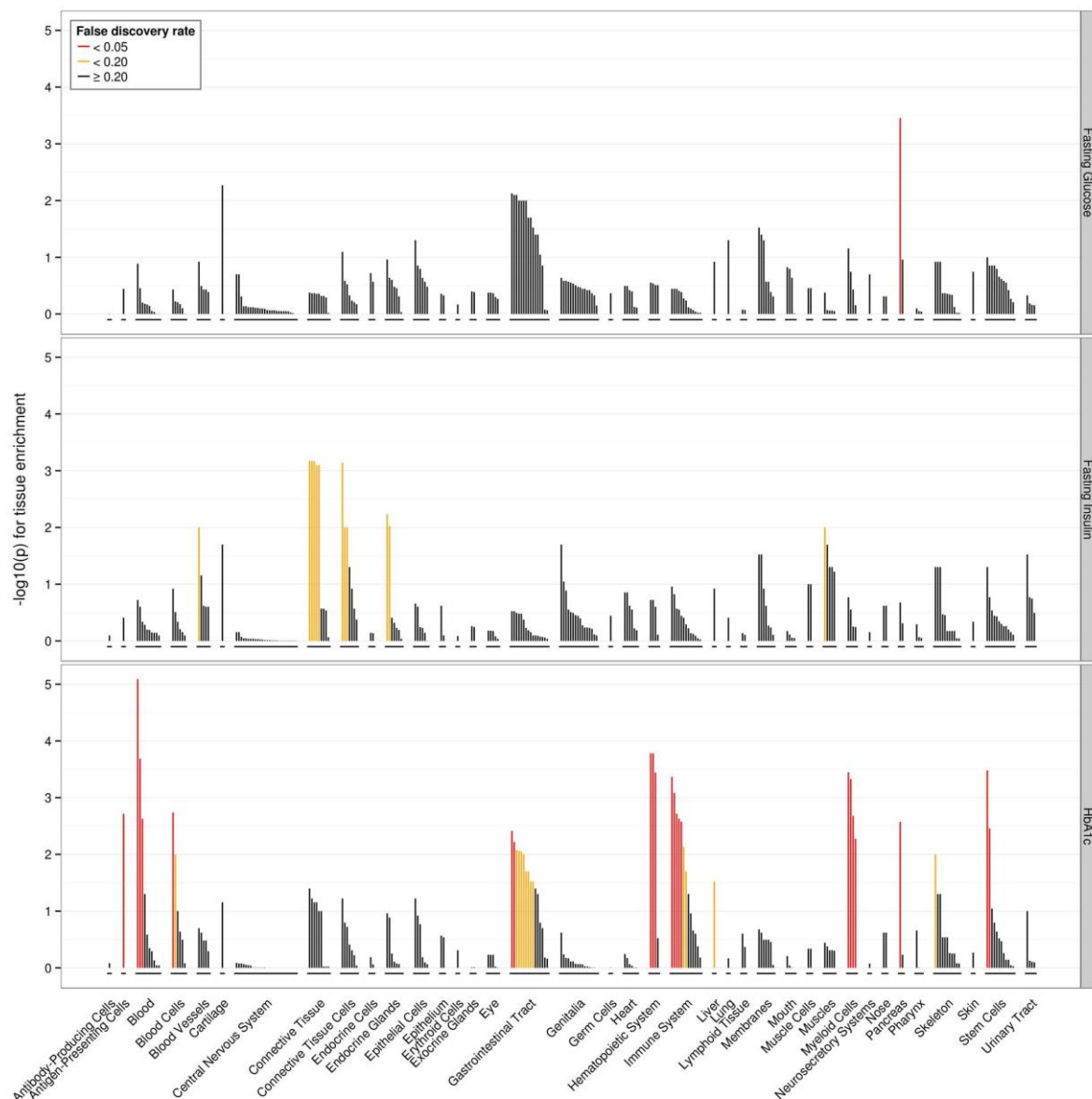


Figure 5 - Tissues and cell types significantly enriched for genes within glycaemic-associated loci.
Top panel FG-associated loci, middle panel FI-associated loci, bottom panel Hba1c-associated loci.
FRD thresholds are shown in red ($q < 0.05$), orange ($q < 0.2$), grey ($q \geq 0.2$).

The association between FI-associated variants and genes expressed in adrenal glands is notable. Incidental adrenal masses (those detected through routine use of radiological imaging but for which patients have not yet shown signs of adrenal hormone excess) have often been associated with hypertension, dyslipidaemia, glucose intolerance, and obesity, all hallmarks of insulin resistance⁶⁵. However, it has not been clear whether incidental adrenal masses are a cause or consequence of the associated insulin resistance^{66,67}. Our results would suggest that FI-associated variants (a surrogate for insulin resistance) are enriched in genes expressed in the adrenal glands, suggesting a possible direct role for these in insulin resistance. One hypothesis is that these genes might influence cortisol levels, which could subsequently contribute to insulin resistance and FI levels through impairment of the insulin receptor signalling pathway in peripheral tissues, as well as influencing body fat distribution, stimulate lipolysis, and other indirect mechanisms^{67,68}.

Gene-set Analyses

Next, we performed gene-set analysis using DEPICT (**Methods**). In keeping with previous results ²⁷, we found distinct gene-sets enriched (FDR<0.05) for each glycaemic trait (except 2hGlu, for which genome-wide associations were insufficient to have power in this analysis). FG-associated variants highlighted gene-sets involved in metabolism in addition to gene-sets involved in more general cellular function such as “cytoplasmic vesicle membrane” and “circadian clock” (Figure 6A). In contrast, in addition to metabolism related gene-sets FI-associated variants highlighted pathways related to growth, cancer and reproduction (Figure 6B). This is consistent with the role of insulin as a mitogenic hormone, and with epidemiological links between insulin and certain types of cancer ⁶⁹ and reproductive disorders such as polycystic ovary syndrome ⁷⁰. HbA1c-associated variants highlighted a wide network of gene-sets (Figure 6C), including those linked to metabolism, as well as those linked to haematopoiesis, again recapitulating our postulated effects of variants on glucose and RBC biology. Additional pathways highlighted from HbA1c-associated variants also highlighted previous “CREBP PPi” and lipid biology related to T2D ⁷¹ and HbA1c ⁷², respectively, and potential new biology through which variants may influence HbA1c.

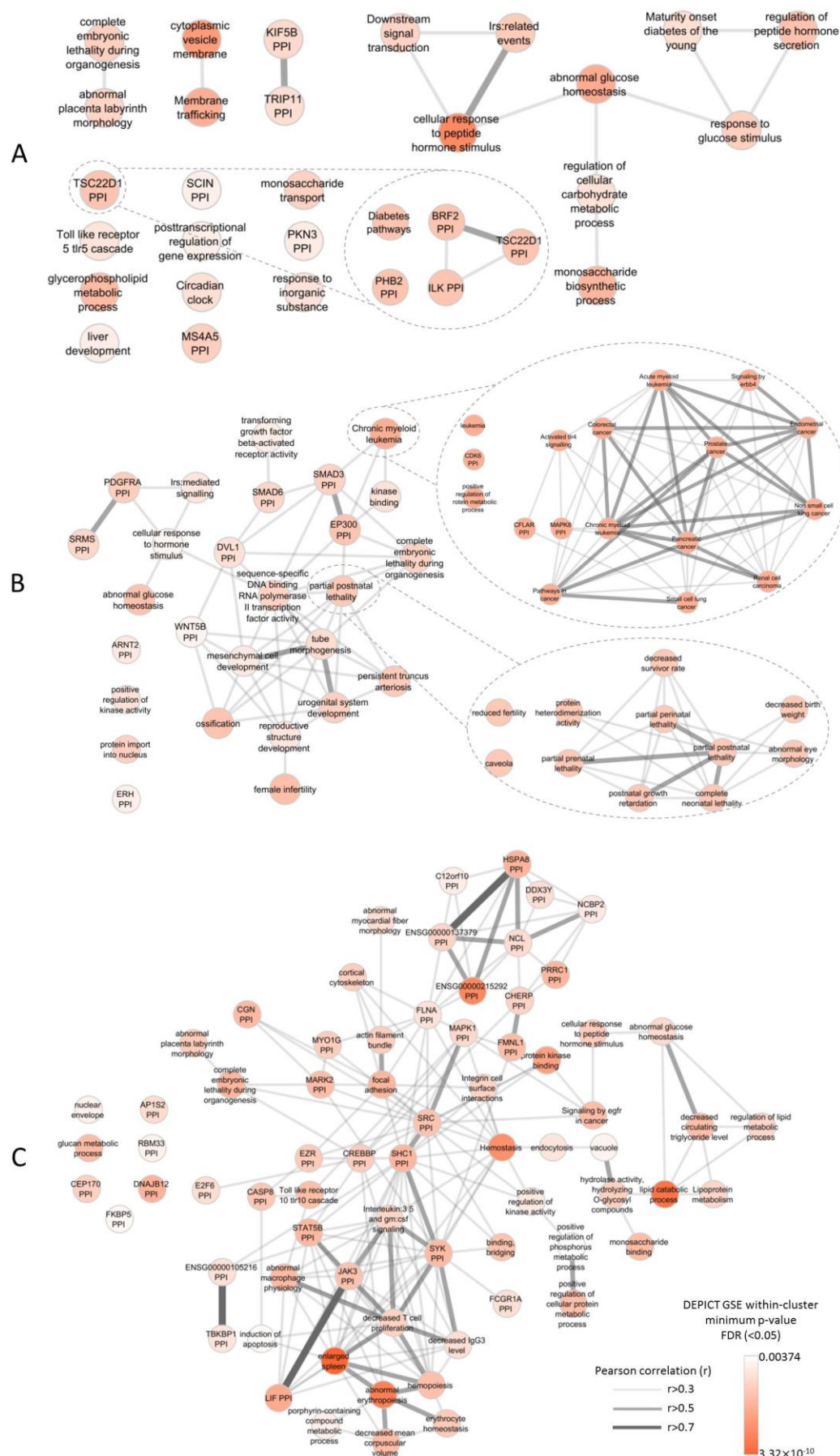


Figure 6 - Gene-set enrichment analyses. Results from affinity-propagation clustering of significantly enriched gene sets (FDR<0.05) identified by DEPICT for A) FG, B) FI, and C) HbA1c. Each node is a cluster of gene-sets represented by an exemplar gene-set with similarities between the clusters represented by the Pearson correlation coefficients ($r>0.3$). The nodes are coloured according to the minimum gene-set enrichment p-value for gene-sets in that cluster. Example clusters are expanded to show the contributing gene-sets.

Discussion

Here we describe a large meta-analysis of GWAS of glycaemic traits for which 30% of the population was composed of East Asian, Hispanic, African-American, South Asian and sub-Saharan African participants, in addition to the European ancestry participants. Overall, this effort identified 242 loci (235 trans-ancestry and seven single-ancestry), which jointly explain between 0.7% (2hGlu in European ancestry individuals, SE=0.85% for 2hGlu) and 6% (HbA1c in African American ancestry, SE=1.2% for HbA1c) of the variance in glycaemic traits in any given ancestry.

A key aim of our study was to evaluate the added advantage of including population diversity into genetic discovery and fine-mapping efforts. Beyond the overall larger sample size included in the trans-ancestry meta-analysis, we were able to estimate the contribution of non-European ancestry data in locus discovery and fine-mapping resolution. We found that 24 of the 99 newly discovered loci owe their discovery to the inclusion of East Asian, Hispanic, African-American, South Asian and sub-Saharan African participant data, due to differences in EAF and effect sizes across ancestries.

Comparison of 295 trans-ancestry lead variants (315 locus-trait associations) across ancestries demonstrated that between 81.5% (for HbA1c) and 85.7% (for FG) of the trans-ancestry lead variants had no evidence of trans-ancestry heterogeneity in allelic effects ($P>0.05$). Expanded analyses including variants across the whole genome, demonstrated at least nominal concordance in the direction of effects between populations of European ancestry and other ancestries for all but the least significant association signals observed in European ancestry GWAS. These observations are consistent with a tail of variants with modest but homogenous effects on glycaemic traits across ancestries that would be amenable to discovery with even larger sample sizes in trans-ancestry meta-analysis.

Recently, there has been ongoing debate regarding the utility of GS for trait prediction across different studies, and in particular, across diverse populations⁷³⁻⁷⁶. Here, we attempted to more precisely estimate the utility of weighted GS constructed from European ancestry effect sizes, to predict trait variance explained in studies independent from those used in genetic discovery, and across ancestries. We show that GSs built from index variants in Europeans ($P<5\times 10^{-8}$) explain similar proportions of the trait variance across populations, though much of the trait variance remains unexplained even in European ancestry individuals. Also, these European participant-based scores will fail to detect ancestry-specific variant effects which can have large effect sizes on specific traits⁷. Consequently, when these analyses are extended to variants with weaker associations, we observe that while (as expected) variance explained is improved in European ancestry participants performance is worsened into other ancestries. We suggest that for the less stringent thresholds, the association signals will be less “peaked” in European ancestry GWAS. As a result, the SNP

selected for the GS is less likely to be causal, meaning that differences in LD structure compared with other ancestry groups adds noise to the prediction.

We further demonstrate that fine-mapping resolution is improved in trans-ancestry, compared to single-ancestry fine-mapping efforts. In ~50% of our loci, we were able to demonstrate the improvement is due to differences in EAF, effect size, or LD structure between ancestries, and not just due to the overall increased sample size available for trans-ancestry fine-mapping. By performing trans-ancestry fine-mapping, and co-localising GWAS signals with eQTL signals and coding variants, we identify new candidate causal genes. Altogether, these results provide additional strong motivation for continued expansion of genetic and genomic efforts in diverse populations, not least to improve understanding of these traits in diverse ancestries in whom individuals are often disproportionately affected by T2D.

Given data on four different glycaemic traits, and their utility to diagnose and monitor T2D and metabolic health, we also sought to characterise biological features underlying these traits. We show that despite significant sharing of genetic loci across the four glycaemic traits, each trait is also characterised by a unique set of features based on stretch enhancer, gene expression and gene-set signatures. Combining genetic data from these traits with T2D data will further elucidate pathways driving normal physiology and pathophysiology, and help further develop useful predictive scores for disease classification and management ^{4,5}.

Acknowledgments

A full list of acknowledgments and funding appears in the **Supplementary Note**.

Methods

Study design and participants

This study included trait data from four glycaemic traits: fasting glucose (FG), fasting insulin (FI), 2hr post-challenge glucose (2hGlu), and glycated haemoglobin (HbA1c). The total number of contributing cohorts ranged from 41 (2hGlu) to 131 (FG), and the maximum sample size for each trait ranged from 85,916 (2hGlu) to 281,416 (FG) (**Supplementary Table 1**). Overall, European ancestry (EUR) participants dominated the sample size for all traits, representing between 68.0% (HbA1c) to 73.8% (2hGlu) of the overall sample size. African Americans (AA) represented between 1.7% (2hGlu) to 5.9% (FG) of participants; individuals of Hispanic ancestry (HISP) represented between 6.8% (FG) to 14.6% (2hGlu) of participants; individuals of East-Asian ancestry (EAS) represented between 9.9% (2hGlu) to 15.4% (HbA1c) of participants; and South-Asian ancestry (SAS) individuals represented between 0% (no contribution to 2hGlu) to 4.4% (HbA1c) of participants. Data from Ugandan participants were only available for the HbA1c analysis and represented 2% of participants.

Phenotypes

Analyses included data for FG and 2hGlu measured in mmol/l, FI measured in pmol/l, and HbA1c in % [where possible, studies reported HbA1c as a National Glycohemoglobin Standardization Program

(NGSP) percent]. Similar to previous MAGIC efforts⁷, individuals were excluded if they had type 1 or type 2 diabetes (defined by physician diagnosis); reported use of diabetes-relevant medication(s); or had a FG ≥ 7 mmol/L, 2hGlu ≥ 11.1 mmol/L, or HbA1c $\geq 6.5\%$, as detailed in **Supplementary Table 1**. 2hGlu measures were obtained 120 minutes after a glucose challenge in an oral glucose tolerance test (OGTT). Measures for FG and FI taken from whole blood were corrected to plasma level using the correction factor 1.13⁷⁷.

Genotyping, quality control, and imputation

Each participating cohort performed study-level quality control, imputation, and association analyses following a shared analysis plan. Cohorts were genotyped using commercially available genome-wide arrays or the Illumina CardioMetabochip (Metabochip) array (**Supplementary Table 1**)⁷⁸. Prior to imputation, each cohort performed stringent sample and variant quality control (QC) to ensure only high-quality variants were kept in the genotype scaffold for imputation. Sample quality control checks included removing samples with low call rate $< 95\%$, extreme heterozygosity, sex mismatch with X chromosome variants, duplicates, first- or second-degree relatives (unless by design), or ancestry outliers. Following sample QC, cohorts applied variant QC thresholds for call rate ($< 95\%$), Hardy-Weinberg Equilibrium (HWE) $P < 1 \times 10^{-6}$, and minor allele frequency (MAF). Full details of QC thresholds and exclusions by participating cohort are available in **Supplementary Table 1**.

Imputation was performed up to the 1000 Genomes Project phase 1 (v3) cosmopolitan reference panel⁷⁹, with a small number of cohorts imputing up to the 1000 Genomes phase 3 panel¹⁹ or population-specific reference panels (**Supplementary Table 1**).

Study level association analyses

Each of the glycaemic traits (FG, natural log FI, and 2hGlu) were regressed on BMI (except HbA1c), study-specific covariates, and principal components (unless implementing a linear mixed model). Analyses for FG, FI, and 2hGlu were adjusted for BMI as we had previously shown this did not materially affect results for FG and 2hGlu but improved our ability to detect FI-associated loci¹⁵. For simplicity, we refer to the traits as FG, FI and 2hGlu. For a discussion on collider bias see **Supplementary Note section 2c**. Both the raw and rank-based inverse normal transformed residuals from the regression were tested for association with genetic variants using SNPTTEST²³ or Mach2QTL^{80,81}. Poorly imputed variants, defined as imputation $r^2 < 0.4$ or INFO score < 0.4 , were excluded from downstream analyses (**Supplementary Table 1**). Following study level QC, approximately 12,229,036 variants (GWAS cohorts) and 1,999,204 variants (Metabochip cohorts) were available for analysis (**Supplementary Table 1**).

Centralised quality control

Each contributing cohort shared their summary statistic results with the central analysis group who performed additional QC using EasyQC⁸². Allele frequency estimates were compared to estimates from 1000Gp1 reference panel⁷⁹, and variants were excluded from downstream analyses if there was a minor allele frequency difference > 0.2 for AA, EUR, HISP, and EAS populations against AFR, EUR, MXL, and ASN populations from 1000 Genomes Phase 1, respectively, or a minor allele frequency difference > 0.4 for SAS against EUR populations. At this stage, additional variants were excluded from each cohort file if they met one of the following criteria: were tri-allelic; had a minor

allele count (MAC) < 3; demonstrated a standard error of the effect size ≥ 10 ; or were missing an effect estimate, standard error, or imputation quality. All data that survived QC (approximately 12,186,053 variants from GWAS cohorts and 1,998,657 variants from Metabochip cohorts) were available for downstream meta-analyses.

Single-ancestry meta-analyses

Single-ancestry meta-analyses were performed within each ancestry group using the fixed-effects inverse variance meta-analysis implemented in METAL²⁰. We applied a double-genomic control (GC) correction^{15,83} to both the study-specific GWAS results and the single-ancestry meta-analysis results. Study-specific Metabochip results were GC-corrected using 4,973 SNPs included on the Metabochip array for replication of associations with QT-interval, a phenotype not correlated with our glycaemic traits¹⁵.

Identification of single-ancestry index variants

To identify distinct association index variants across each chromosome within each ancestry (**Glossary box**), we performed approximate conditional analyses implemented in GCTA²¹ using the --cojo-slc option (autosomes) and distance-based clumping (X chromosome). Linkage disequilibrium (LD) correlations for GCTA were estimated from a representative cohort from each ancestry: WGHS (EUR); CHNS (EAS); SINDI (SAS); BioMe (AA); SOL (HISP) and Uganda (for itself). The results from GCTA were comparable when using alternative cohorts for the LD reference. For any index variant with a QC flag which caused reason for concern, we performed manual inspection of forest plots to decide whether the signal was likely to be real (**Supplementary note**). Among 335 single-ancestry index variants across all traits, this manual inspection was done for 40 signals of which 32 passed and 8 failed after inspection. Thus, a total of 327 single-ancestry index variants passed and 8 failed.

Trans-ancestry meta-analyses

To leverage power across all ancestries, we also conducted trait-specific trans-ancestry meta-analysis by combining the single-ancestry meta-analysis results using MANTRA (**Supplementary Figure 3**)⁸⁴. We defined \log_{10} Bayes' Factor (BF) > 6 as genome-wide significant, approximately comparable to $P < 5 \times 10^{-8}$.

Manual curation of trans-ancestry lead variants

To ensure trans-ancestry lead variants were robust, we performed manual inspection of forest plots by at least two authors, for any variants with flags indicating possible QC issues (**Supplementary Note**). Of 463 trans-ancestry lead variants across all traits, 184 passed without inspection, 131 passed after inspection, and 148 failed after inspection.

Correlation in EAF and heterogeneity in effect sizes of TA lead variants across ancestries

For each pair of ancestries, we calculated Pearson's correlation in EAFs for each trans-ancestry lead variant. The pairwise summarised heterogeneity of effect sizes between ancestries was then tested using the joint F-test of heterogeneity²⁹. The test statistic is the sum of Cochran Q-statistics for heterogeneity across all trans-ancestry signals. Under the null hypothesis, the statistics follows the χ^2 distribution with n degrees of freedom, where n is the number of the trans-ancestry lead variants.

Concordance analyses of LD pruned European single-ancestry index variants into other ancestries

We compared the direction of effect of variants on each trait separately. For each trait, we identified variants reported in the European ancestry meta-analysis and each non-European ancestry meta-analysis, in turn. These variants were assigned to P -value bins, according to the strength of the association with the trait in the European ancestry meta-analysis: $P < 5 \times 10^{-8}$; $5 \times 10^{-8} \leq P < 5 \times 10^{-6}$; $5 \times 10^{-6} \leq P < 5 \times 10^{-4}$; $5 \times 10^{-4} \leq P < 0.05$; and $P \geq 0.05$. Within each P -value bin, we selected a set of “independent” variants that were separated by 1 Mb. We defined independence using a distance-based threshold because of differences in patterns of LD between ancestry groups. For each P -value bin, the proportion of variants with the same direction of effect on the trait between the two ancestries was calculated along with a P -value from the binomial test to determine if the proportion of variants with the same direction of effect was greater than that expected by chance (50%, one sided).

LD-pruned variant lists

Several downstream analyses (for example, genomic feature enrichment, genetic scores, and estimation of variance explained by associated variants) require independent LD-pruned variants ($r^2 < 0.1$) to avoid double-counting variants which might otherwise be in LD with each other and that do not provide additional “independent” evidence. Therefore, for these analyses we generated different lists of either TA or single-ancestry LD pruned ($r^2 < 0.1$) variants, keeping in each case the variant with the strongest evidence of association (**Supplementary Table 7**). Subsequently, we combined TA and single-ancestry variant lists and conducted further LD pruning. For some analyses, we took the TA pruned variant list and added single-ancestry signals if the LD $r^2 < 0.1$, while for others we started with the single-ancestry pruned lists and supplemented with TA lead variants if the LD $r^2 < 0.1$. One exception was the list used for eQTL co-localisations, which included all single-ancestry European signals (without LD pruning) and supplemented with any additional TA lead variants (starting from the variants with the most significant P -values) in EUR LD $r^2 < 0.1$ with any of the variants already in list, and that reached at least $P < 1 \times 10^{-5}$ in the European ancestry meta-analysis.

Transferability of genetic scores (GSs) across ancestries

To determine the power of a European-based genetic score (GS) to predict trait values within non-European populations, we began with the list of European LD-pruned index variants (**Supplementary Table 7**) and their effect sizes. We first tested the GS in four European cohorts with individual level data that did not contribute to this meta-analysis: WHITEHALL II, Oxford Biobank, VIKING and UKHLS (**Supplementary Table 1**). We used individual level genotype data to build an effect-size weighted GS for each individual, and then obtained the trait variance explained via linear regression. We then tested the European GS in each European ancestry cohort contributing to the meta-analysis with $> 1,000$ samples by: (i) adjusting the effect size estimate of each variant to remove the contribution of the cohort³⁰; and (ii) obtaining the proportion of the trait explained by the GS (R^2) on the basis of cohort-level summary statistics using the R package “gtx”. Finally, we obtained the proportion of the trait explained by the European GS in other ancestry groups on the basis of single-ancestry meta-analysis summary statistics using the R package “gtx”. Variants reported in $< 50\%$ of the total sample size in each ancestry group for each trait were excluded as they can yield unstable estimates of R^2 . Standard errors of effect size estimates were adjusted to account for differences in the sample size reported for each variant. Additional sensitivity analyses were also performed using: (i) variants that exhibited only modest evidence of heterogeneity ($P > 1 \times 10^{-6}$) between ancestries from the trans-

ancestry analysis; (ii) variants with no evidence of statistically significant between-ancestry heterogeneity ($P > 0.05$) (iii) variants with $P < 1 \times 10^{-5}$ in the European meta-analysis (1Mb distance clumped); and (iv) variants with $P < 0.05$ in the European meta-analysis (1Mb distance clumped).

Trait variance explained

To determine how much of the phenotypic variance of each trait could be explained by the corresponding trait-associated loci, variants were combined in a series of weighted GS. The analysis was performed in a subset of the cohorts included in the discovery GWAS (with representation from each ancestry) and in a smaller number of independent cohorts (European ancestry only). Up to three GS were generated per trait (and for each ancestry), representing single-ancestry signals, single-ancestry signals plus trans-ancestry signals, and trans-ancestry signals plus single-ancestry signals (**Supplementary Table 7**). In the case of the European ancestry cohorts that contributed to the GWAS, we employed the method of Nolte *et al.*³⁰ to adjust the effect sizes (betas) from the GWAS for the contribution of that cohort, providing sets of cohort-specific effect sizes that were then used to generate the GS. The association between each GS and its corresponding trait was tested by linear regression and the adjusted R^2 from the model extracted as an estimate of the variance explained.

Fine-mapping

Of the 242 loci identified in this study, 237 were autosomal loci which we took forward for fine-mapping (**Supplementary Table 2**). We used the Bayesian fine-mapping method FINEMAP⁸⁵ (version 1.1) to refine association signals and attempt to identify likely causal variants at each locus. FINEMAP estimates the maximum number of causal variants at each locus, calculates the posterior probability of each variant being causal, and proposes the most likely configuration of causal variants. The posterior probabilities of the configurations in each locus were used to construct 99% credible sets.

We performed both single-ancestry and trans-ancestry fine-mapping. In both analyses, only data from cohorts genotyped on GWAS arrays were used, and analyses were limited to trans-ancestry lead variants and other single-ancestry lead variants present in at least 90% of the samples for each trait. For the single-ancestry fine-mapping, FINEMAP estimates the number of causal variants in a region up to a maximum number, which we set to be two plus the number of distinct signals identified from the GCTA signal selection. FINEMAP uses single-ancestry and trait-specific z-scores from the fixed-effect meta-analysis in METAL²⁰ and an ancestry-specific LD reference, which we created from a subset of cohorts (combined sample size > 30% of the sample size for that ancestry), weighting each cohort by sample size. In the trans-ancestry fine-mapping, FINEMAP was similarly used to estimate the number of causal variants starting with two, and trait-specific z-scores and LD maps were generated from the sample size weighted average of those used in the single-ancestry fine-mapping. The maximum number of causal variants was iteratively increased by one until it was larger than the number of causal variants supported by data (Bayes factor), which was the estimated maximum number of causal variants used in the final run of fine-mapping analysis.

To compare fine-mapping results obtained from the single-ancestry and trans-ancestry efforts, analyses were limited to fine-mapping regions with evidence for a single likely causal variant in both, enabling a straightforward comparison of credible sets (**Supplementary note**). To ensure any

difference in the fine-mapping results was not driven by different sets of variants being present in the different analyses, we repeated the single-ancestry fine-mapping limited to the same set of variants used in the trans-ancestry fine-mapping. The fine-mapping resolution was assessed based on comparisons of the 99% credible sets in terms of number of variants included in the set, and length of the region. To assess whether the improvement in the trans-ancestry fine-mapping was due to differences in LD, increased sample size, or both, we repeated the trans-ancestry fine-mapping mimicking the sample size present in the single-ancestry fine-mapping by dividing the standard errors by the square root of the sample size ratio and compared the results with those from the single-ancestry fine-mapping.

Functional Annotation of trait-associated variants

HbA1c signal classification

There were 202 autosomal HbA1c-associated signals from either the single-ancestry (i.e. all GCTA-signals from any ancestry) or trans-ancestry meta-analyses. To classify these signals in terms of their likely mode of action (i.e., glycaemic, erythrocytic, or other⁷), we examined association summary statistics for the lead variants at the 202 signals in other large European datasets for 19 additional traits: three glycaemic traits from this study (FG, 2hGlu and FI); seven mature red blood cell (RBC) traits⁸⁶ (red blood cell count, mean corpuscular volume, haematocrit, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, haemoglobin concentration and red cell distribution width); five reticulocyte traits (reticulocyte count, reticulocyte fraction of red cells, immature fraction of reticulocytes, high light scatter reticulocyte count and high light scatter percentage of red cells)⁸⁶, and four iron traits (serum iron, transferrin, transferrin saturation and ferritin)⁸⁷. Of the 202 autosomal HbA1c signals, data were available for the lead (n=177) or proxy (European LD $r^2 > 0.8$, n = 9) variants at 186 signals.

The additional traits were clustered using hierarchical clustering to ensure biologically related traits would cluster together (**Supplementary note**). We then used a non-negative matrix factorization (NMF)⁸⁸ process to cluster the HbA1c signals. Each cluster was labelled as glycaemic, reticulocyte, mature RBC, or iron related based on the strength of association of signals in the cluster to the glycaemic, reticulocyte, mature RBC and iron traits (**Supplementary note**). To verify that our cluster naming was correct, we used HbA1c association results conditioned on either FG or iron traits, or type 2 diabetes association results (**Supplementary note**).

HbA1c genetic risk scores (GRSs) and type 2 diabetes (T2D) risk

We constructed GRS for each cluster of HbA1c-associated signals (based on hard clustering) and tested the association of each cluster with T2D risk using samples from the UK Biobank. Pairs of HbA1c signals in LD (EUR $r^2 > 0.10$) were LD pruned by removing the signal with the less significant *P*-value of association with HbA1c. The GRS for each cluster was calculated based on the logarithm of odds ratios from the latest T2D study summary statistics⁸⁹ and UK Biobank genotypes imputed to the Haplotype Reference Consortium¹⁹. From 487,409 UK Biobank samples, we excluded participants for the following reasons: 373 with mismatched sex; 9 not used in the kinship calculation; 78,365 non-European ancestry individuals; and 138,504 with missing T2D status, age, or sex information. We further removed 26,896 related participants (kinship > 0.088, preferentially removing individuals with the largest number of relatives and controls where a T2D case was related

to a control). T2D cases were defined by: (i) a history of diabetes without metformin or insulin treatment, (ii) self-reported diagnosis of T2D, or (iii) diagnosis of T2D in a national registry (N = 17,022). Controls were participants without a history of T2D (N = 226,240). We tested for association between each GRS and T2D using logistic regression including covariates for age, sex, and the first five principle components. Significance of association was evaluated by a bootstrap approach to incorporate the variance of each HbA1c associated signal in the T2D summary data. To do this, we generated the GRS of each cluster 200 times by resampling the logarithm of odds ratio of each signal with T2D. For each non-glycaemic class that had a GRS significantly associated with T2D, we performed sensitivity analyses to evaluate whether the association was driven from variants that also belonged to a glycaemic cluster when using a soft clustering approach (the signals were classified as also glycaemic in the soft clustering or had an association $P \leq 0.05$ with any of the three glycaemic traits).

Chromatin states

To identify genetic variants within association signals that overlapped predicted chromatin states, we used a previously published, 13 chromatin state model that included 31 diverse tissues, including pancreatic islets, skeletal muscle, adipose, and liver³³. Briefly, this model was generated from cell/tissue ChIP-seq data for H3K27ac, H3K27me3, H3K36me3, H3K4me1, and H3K4me3, and input control from a diverse set of publicly available data^{47,51,90,91} using the ChromHMM program⁹². As reported previously³³, stretch enhancers were defined as contiguous enhancer chromatin state (Active Enhancer 1 and 2, Genic Enhancer and Weak Enhancer) segments longer than 3kb⁵¹.

Enrichment of genetic variants in genomic features

We used GREGOR (version 1.2.1) to calculate the enrichment of GWAS variants overlapping static and stretch enhancers⁵⁰. For calculating the enrichment of glycaemic trait-associated variants in these annotations, we used the filtered list of trait-associated variants as described above (**Supplementary Table 7**) as input. For calculating the enrichment of sub-classified HbA1c variants, we included the list of loci characterized as Glycaemic, another list of loci characterized as Reticulocyte or mature Red Blood Cell, collectively representing the red blood cell fraction, along with lists of iron related or unclassified loci (**Supplementary Table 17**). We used the following parameters in GREGOR enrichment analyses: European r^2 threshold (for inclusion of variants in LD with the lead variant) = 0.8, LD window size = 1 Mb, and minimum neighbour number = 500.

We used fGWAS (version 0.3.6)⁵² to calculate enrichment of glycaemic trait-associated variants in static and stretch enhancer annotations using summary level GWAS results. We used the default fGWAS parameters for enrichment analyses for individual annotations for each trait. For each annotation, the model provided the natural log of maximum likelihood estimate of the enrichment parameter. Annotations were considered as significantly enriched if the log2 (parameter estimate) and respective 95% confidence intervals were above zero or significantly depleted if the log2 (parameter estimate) and respective 95% confidence intervals were below zero.

We tested enrichment of trait-associated variants in static and stretch enhancer annotations with GARFIELD (v2)⁵³. We formatted annotation overlap files as required by the tool; prepared input data at two GWAS thresholds - of 1×10^{-5} and a more stringent 1×10^{-8} by pruning and clumping with default parameters (garfield-prep-chr script). We calculated enrichment in each individual annotation using

garfield-test.R with $-c$ option set to 0. We also calculated the effective number of annotations using the garfield-Meff-Padj.R script. We used the effective number of annotations for each trait to obtain Bonferroni corrected significance thresholds for enrichment for each trait.

eQTL analyses

To aid in the identification of candidate casual genes at the European-only and trans-ancestry association signals, we examined whether any of the lead variants associated with glycaemic traits (**Supplementary Table 7**) were also associated with expression level (FDR < 5%) of nearby transcripts located within 1 Mb in existing eQTL data sets of blood, subcutaneous adipose, visceral adipose, skeletal muscle, and pancreatic islet samples^{54,55,93-96}. LD was estimated from the collected cohort pairwise LD information, where available, else from the European samples in 1000G phase 3. GWAS and eQTL signals likely co-localise when the GWAS variant and the variant most strongly associated with the expression level of the corresponding transcript (eSNP) exhibit high pairwise LD ($r^2 > 0.8$; 1000 Genomes Phase 3, EUR). At these signals, we conducted reciprocal conditional analyses to test association between the GWAS variant and transcript level when the eSNP was also included in the model, and vice versa. We report GWAS and eQTL signals as co-localised if the association for the eSNP was not significant (FDR \geq 5%) when conditioned on the GWAS variant; we also report signals from the eQTLGen whole blood meta-analysis data that meet only the LD threshold because conditional analysis was not possible.

Tissue and gene-set analysis

We performed enrichment analysis using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) version 3, specifically developed for 1000 Genomes Project imputed meta-analysis data⁹⁷ to identify cell types and tissues in which genes at trait-associated variants were strongly expressed, and to detect enrichment of gene-sets or pathways. DEPICT data included human gene expression data for 19,987 genes in 10,968 reconstituted gene sets, and 209 tissues/cell types. Because gene expression data in DEPICT is based on European samples and LD, we selected trait-associated variants with $P < 10^{-5}$ in the European meta-analysis and tested for enrichment of signals in each reconstituted gene-set, and each tissue or cell type. Enrichment results with a false discovery rate (FDR) < 0.05 were considered significant. We ran DEPICT based on association results for all traits among: (i) cohorts with genome-wide data, or (ii) all cohorts (genome-wide and Metabochip cohorts). Because results were broadly consistent between the two approaches, we present results from the analysis that contained all cohorts as it had greater statistical power.

References

- 1 in *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation* (World Health Organization Copyright © World Health Organization 2011., 2011).
- 2 Goodarzi, M. O. *et al.* Fasting insulin reflects heterogeneous physiological processes: role of insulin clearance. *American journal of physiology. Endocrinology and metabolism* **301**, E402-408, doi:10.1152/ajpendo.00013.2011 (2011).
- 3 Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-2171, doi:10.2337/db13-0949 (2014).

- 1436 4 Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations
1437 point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS medicine*
1438 **15**, e1002654, doi:10.1371/journal.pmed.1002654 (2018).
- 1439 5 Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for
1440 Diabetes Diagnosis and Precision Medicine. *Endocrine reviews* **40**, 1500-1520,
1441 doi:10.1210/er.2019-00088 (2019).
- 1442 6 Sarwar, N. *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of
1443 vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*
1444 **375**, 2215-2222, doi:10.1016/s0140-6736(10)60484-9 (2010).
- 1445 7 Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on
1446 type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic
1447 genome-wide meta-analysis. *PLoS medicine* **14**, e1002383,
1448 doi:10.1371/journal.pmed.1002383 (2017).
- 1449 8 Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their
1450 impact on type 2 diabetes risk. *Nature genetics* **42**, 105-116, doi:10.1038/ng.520
1451 (2010).
- 1452 9 Manning, A. K. *et al.* A genome-wide approach accounting for body mass index
1453 identifies genetic variants influencing fasting glycemic traits and insulin resistance.
1454 *Nature genetics* **44**, 659-669, doi:10.1038/ng.2274 (2012).
- 1455 10 Walford, G. A. *et al.* Genome-Wide Association Study of the Modified Stumvoll
1456 Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity
1457 Loci. *Diabetes* **65**, 3200-3211, doi:10.2337/db16-0199 (2016).
- 1458 11 Horikoshi, M. *et al.* Discovery and Fine-Mapping of Glycaemic and Obesity-Related
1459 Trait Loci Using High-Density Imputation. *PLoS genetics* **11**, e1005230,
1460 doi:10.1371/journal.pgen.1005230 (2015).
- 1461 12 Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding
1462 variants influencing glycemic traits define an effector transcript at the G6PC2-
1463 ABCB11 locus. *PLoS genetics* **11**, e1004876, doi:10.1371/journal.pgen.1004876
1464 (2015).
- 1465 13 Hwang, J. Y. *et al.* Genome-wide association meta-analysis identifies novel variants
1466 associated with fasting plasma glucose in East Asians. *Diabetes* **64**, 291-298,
1467 doi:10.2337/db14-0563 (2015).
- 1468 14 Chen, P. *et al.* Multiple nonglycemic genomic loci are newly associated with blood
1469 level of glycated hemoglobin in East Asians. *Diabetes* **63**, 2551-2562,
1470 doi:10.2337/db13-1815 (2014).
- 1471 15 Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing
1472 glycemic traits and provide insight into the underlying biological pathways. *Nature*
1473 *genetics* **44**, 991-1005, doi:10.1038/ng.2385 (2012).
- 1474 16 Spanakis, E. K. & Golden, S. H. Race/ethnic difference in diabetes and diabetic
1475 complications. *Current diabetes reports* **13**, 814-823, doi:10.1007/s11892-013-0421-
1476 9 (2013).
- 1477 17 Tillin, T. *et al.* Insulin resistance and truncal obesity as important determinants of the
1478 greater incidence of diabetes in Indian Asians and African Caribbeans compared with
1479 Europeans: the Southall And Brent REvisited (SABRE) cohort. *Diabetes care* **36**, 383-
1480 393, doi:10.2337/dc12-0544 (2013).

- 1481 18 Whincup, P. H. *et al.* Early emergence of ethnic differences in type 2 diabetes
1482 precursors in the UK: the Child Heart and Health Study in England (CHASE Study).
1483 *PLoS medicine* **7**, e1000263, doi:10.1371/journal.pmed.1000263 (2010).
- 1484 19 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74,
1485 doi:10.1038/nature15393 (2015).
- 1486 20 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
1487 genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190-2191,
1488 doi:10.1093/bioinformatics/btq340 (2010).
- 1489 21 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
1490 complex trait analysis. *American journal of human genetics* **88**, 76-82,
1491 doi:10.1016/j.ajhg.2010.11.011 (2011).
- 1492 22 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics
1493 identifies additional variants influencing complex traits. *Nature genetics* **44**, 369-375,
1494 s361-363, doi:10.1038/ng.2213 (2012).
- 1495 23 Genome-wide association study of 14,000 cases of seven common diseases and
1496 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).
- 1497 24 Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution
1498 using high-density imputation and islet-specific epigenome maps. *Nature genetics*
1499 **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).
- 1500 25 Luo, Y. *et al.* Transcription factor Ets1 regulates expression of thioredoxin-interacting
1501 protein and inhibits insulin secretion in pancreatic beta-cells. *PloS one* **9**, e99049,
1502 doi:10.1371/journal.pone.0099049 (2014).
- 1503 26 Braccini, L. *et al.* PI3K-C2gamma is a Rab5 effector selectively controlling endosomal
1504 Akt2 activation downstream of insulin signalling. *Nature communications* **6**, 7400,
1505 doi:10.1038/ncomms8400 (2015).
- 1506 27 Ng, N. H. J. *et al.* Tissue-Specific Alteration of Metabolic Pathways Influences
1507 Glycemic Regulation. *bioRxiv*, 790618, doi:10.1101/790618 (2019).
- 1508 28 Aschard, H., Vilhjalmsen, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for
1509 heritable covariates can bias effect estimates in genome-wide association studies.
1510 *American journal of human genetics* **96**, 329-339, doi:10.1016/j.ajhg.2014.12.021
1511 (2015).
- 1512 29 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide
1513 association study of educational attainment in 1.1 million individuals. *Nature*
1514 *genetics* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).
- 1515 30 Nolte, I. M. *et al.* Missing heritability: is the gap closing? An analysis of 32 complex
1516 traits in the Lifelines Cohort Study. *European journal of human genetics : EJHG* **25**,
1517 877-885, doi:10.1038/ejhg.2017.50 (2017).
- 1518 31 Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation defines causal
1519 mechanisms at type 2 diabetes susceptibility loci. *Nature genetics* **47**, 1415-1425,
1520 doi:10.1038/ng.3437 (2015).
- 1521 32 Spracklen, C. N. *et al.* Identification and functional analysis of glycemic trait loci in
1522 the China Health and Nutrition Survey. *PLoS genetics* **14**, e1007275,
1523 doi:10.1371/journal.pgen.1007275 (2018).
- 1524 33 Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression
1525 and type 2 diabetes. *Proceedings of the National Academy of Sciences of the United*
1526 *States of America* **114**, 2301-2306, doi:10.1073/pnas.1621192114 (2017).

1527 34 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS
1528 Power. *American journal of human genetics* **104**, 65-75,
1529 doi:10.1016/j.ajhg.2018.11.008 (2019).

1530 35 Shriner, D. & Rotimi, C. N. Whole-Genome-Sequence-Based Haplotypes Reveal Single
1531 Origin of the Sickle Allele during the Holocene Wet Phase. *American journal of*
1532 *human genetics* **102**, 547-556, doi:10.1016/j.ajhg.2018.02.003 (2018).

1533 36 Kramer, H. J. *et al.* African Ancestry-Specific Alleles and Kidney Disease Risk in
1534 Hispanics/Latinos. *Journal of the American Society of Nephrology : JASN* **28**, 915-922,
1535 doi:10.1681/asn.2016030357 (2017).

1536 37 Ravenhall, M. *et al.* Novel genetic polymorphisms associated with severe malaria and
1537 under selective pressure in North-eastern Tanzania. *PLoS genetics* **14**, e1007172,
1538 doi:10.1371/journal.pgen.1007172 (2018).

1539 38 Hodonsky, C. J. *et al.* Genome-wide association study of red blood cell traits in
1540 Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS*
1541 *genetics* **13**, e1006760, doi:10.1371/journal.pgen.1006760 (2017).

1542 39 Gurdasani, D. *et al.* Uganda Genome Resource Enables Insights into Population
1543 History and Genomic Discovery in Africa. *Cell* **179**, 984-1002.e1036,
1544 doi:10.1016/j.cell.2019.10.004 (2019).

1545 40 Rees, M. G. *et al.* Cellular characterisation of the GCKR P446L variant associated with
1546 type 2 diabetes risk. *Diabetologia* **55**, 114-122, doi:10.1007/s00125-011-2348-5
1547 (2012).

1548 41 Bonomo, J. A. *et al.* The ras responsive transcription factor RREB1 is a novel
1549 candidate gene for type 2 diabetes associated end-stage kidney disease. *Human*
1550 *molecular genetics* **23**, 6441-6447, doi:10.1093/hmg/ddu362 (2014).

1551 42 Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting
1552 glucose and type 2 diabetes susceptibility. *Nature communications* **6**, 5897,
1553 doi:10.1038/ncomms6897 (2015).

1554 43 Scott, R. A. *et al.* A genomic approach to therapeutic target validation identifies a
1555 glucose-lowering GLP1R variant protective for coronary heart disease. *Science*
1556 *translational medicine* **8**, 341ra376, doi:10.1126/scitranslmed.aad3744 (2016).

1557 44 Nai, A. *et al.* TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum
1558 hepcidin levels in normal individuals. *Blood* **118**, 4459-4462, doi:10.1182/blood-
1559 2011-06-364034 (2011).

1560 45 Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C)
1561 levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-3239,
1562 doi:10.2337/db10-0502 (2010).

1563 46 Sarnowski, C. *et al.* Impact of Rare and Common Genetic Variants on Diabetes
1564 Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for
1565 Precision Medicine Program. *American journal of human genetics* **105**, 706-718,
1566 doi:10.1016/j.ajhg.2019.08.010 (2019).

1567 47 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*
1568 **518**, 317-330, doi:10.1038/nature14248 (2015).

1569 48 Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in
1570 449,484 individuals identifies novel genetic loci and pathways. *Nature genetics* **50**,
1571 920-927, doi:10.1038/s41588-018-0151-7 (2018).

- 1572 49 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals
1573 identifies new genetic and functional links to intelligence. *Nature genetics* **50**, 912-
1574 919, doi:10.1038/s41588-018-0152-6 (2018).
- 1575 50 Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated
1576 variants in epigenomic features using a systematic, data-driven approach.
1577 *Bioinformatics (Oxford, England)* **31**, 2601-2606, doi:10.1093/bioinformatics/btv201
1578 (2015).
- 1579 51 Parker, S. C. *et al.* Chromatin stretch enhancer states drive cell-specific gene
1580 regulation and harbor human disease risk variants. *Proceedings of the National*
1581 *Academy of Sciences of the United States of America* **110**, 17921-17926,
1582 doi:10.1073/pnas.1317023110 (2013).
- 1583 52 Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association
1584 studies of 18 human traits. *American journal of human genetics* **94**, 559-573,
1585 doi:10.1016/j.ajhg.2014.03.004 (2014).
- 1586 53 Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through
1587 integration of functional annotations with association signals. *Nature genetics* **51**,
1588 343-353, doi:10.1038/s41588-018-0322-6 (2019).
- 1589 54 van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory
1590 Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic
1591 Traits to Their Downstream Effectors. *PLoS genetics* **11**, e1005694,
1592 doi:10.1371/journal.pgen.1005694 (2015).
- 1593 55 Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-
1594 Metabolic Traits. *American journal of human genetics* **100**, 428-443,
1595 doi:10.1016/j.ajhg.2017.01.027 (2017).
- 1596 56 Scott, L. J. *et al.* The genetic regulatory signature of type 2 diabetes in human
1597 skeletal muscle. *Nature communications* **7**, 11764, doi:10.1038/ncomms11764
1598 (2016).
- 1599 57 Ben Harouch, S., Klar, A. & Falik Zaccai, T. C. in *GeneReviews((R))* (eds M. P. Adam *et al.*)
1600 (University of Washington, Seattle
1601 University of Washington, Seattle. GeneReviews is a registered trademark of the University
1602 of Washington, Seattle. All rights reserved., 1993).
- 1603 58 Agus, D. B. *et al.* Vitamin C crosses the blood-brain barrier in the oxidized form
1604 through the glucose transporters. *The Journal of clinical investigation* **100**, 2842-
1605 2848, doi:10.1172/jci119832 (1997).
- 1606 59 Wolking, S. *et al.* Focal epilepsy in glucose transporter type 1 (Glut1) defects: case
1607 reports and a review of literature. *Journal of neurology* **261**, 1881-1886,
1608 doi:10.1007/s00415-014-7433-5 (2014).
- 1609 60 Guallar, D. *et al.* RNA-dependent chromatin targeting of TET2 for endogenous
1610 retrovirus control in pluripotent stem cells. *Nature genetics* **50**, 443-451,
1611 doi:10.1038/s41588-018-0060-9 (2018).
- 1612 61 Bian, F. *et al.* TET2 facilitates PPARgamma agonist-mediated gene regulation and
1613 insulin sensitization in adipocytes. *Metabolism: clinical and experimental* **89**, 39-47,
1614 doi:10.1016/j.metabol.2018.08.006 (2018).
- 1615 62 Yoo, Y. *et al.* TET-mediated hydroxymethylcytosine at the Ppargamma locus is
1616 required for initiation of adipogenic differentiation. *International journal of obesity*
1617 (2005) **41**, 652-659, doi:10.1038/ijo.2017.8 (2017).

1618 63 Lees, J. A. *et al.* Lipid transport by TMEM24 at ER-plasma membrane contacts
1619 regulates pulsatile insulin secretion. *Science (New York, N.Y.)* **355**,
1620 doi:10.1126/science.aah6171 (2017).

1621 64 Pottekat, A. *et al.* Insulin biosynthetic interaction network component, TMEM24,
1622 facilitates insulin reserve pool release. *Cell reports* **4**, 921-930,
1623 doi:10.1016/j.celrep.2013.07.050 (2013).

1624 65 Terzolo, M. *et al.* Adrenal incidentaloma: a new cause of the metabolic syndrome?
1625 *The Journal of clinical endocrinology and metabolism* **87**, 998-1003,
1626 doi:10.1210/jcem.87.3.8277 (2002).

1627 66 Muscogiuri, G., Colao, A. & Orio, F. Insulin-Mediated Diseases: Adrenal Mass and
1628 Polycystic Ovary Syndrome. *Trends in endocrinology and metabolism: TEM* **26**, 512-
1629 514, doi:10.1016/j.tem.2015.07.010 (2015).

1630 67 Altieri, B. *et al.* Adrenocortical tumors and insulin resistance: What is the first step?
1631 *International journal of cancer* **138**, 2785-2794, doi:10.1002/ijc.29950 (2016).

1632 68 Androulakis, I. *et al.* Patients with apparently nonfunctioning adrenal incidentalomas
1633 may be at increased cardiovascular risk due to excessive cortisol secretion. *The*
1634 *Journal of clinical endocrinology and metabolism* **99**, 2754-2762,
1635 doi:10.1210/jc.2013-4064 (2014).

1636 69 Johansson, M. *et al.* The influence of obesity-related factors in the etiology of renal
1637 cell carcinoma-A mendelian randomization study. *PLoS medicine* **16**, e1002724,
1638 doi:10.1371/journal.pmed.1002724 (2019).

1639 70 Diamanti-Kandarakis, E. & Dunaif, A. Insulin resistance and the polycystic ovary
1640 syndrome revisited: an update on mechanisms and implications. *Endocrine reviews*
1641 **33**, 981-1030, doi:10.1210/er.2011-1034 (2012).

1642 71 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic
1643 architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981-990,
1644 doi:10.1038/ng.2383 (2012).

1645 72 Leong, A. *et al.* Mendelian Randomization Analysis of Hemoglobin A(1c) as a Risk
1646 Factor for Coronary Artery Disease. *Diabetes care* **42**, 1202-1208, doi:10.2337/dc18-
1647 1712 (2019).

1648 73 Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk
1649 scores. *Human molecular genetics* **28**, R133-r142, doi:10.1093/hmg/ddz187 (2019).

1650 74 Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse
1651 human populations. *Nature communications* **10**, 3328, doi:10.1038/s41467-019-
1652 11112-0 (2019).

1653 75 Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth
1654 to Adulthood. *Cell* **177**, 587-596.e589, doi:10.1016/j.cell.2019.03.028 (2019).

1655 76 Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an
1656 ancestry group. *eLife* **9**, doi:10.7554/eLife.48376 (2020).

1657 77 D'Orazio, P. *et al.* Approved IFCC recommendation on reporting results for blood
1658 glucose (abbreviated). *Clinical chemistry* **51**, 1573-1576,
1659 doi:10.1373/clinchem.2005.051979 (2005).

1660 78 Voight, B. F. *et al.* The metabochip, a custom genotyping array for genetic studies of
1661 metabolic, cardiovascular, and anthropometric traits. *PLoS genetics* **8**, e1002793,
1662 doi:10.1371/journal.pgen.1002793 (2012).

1663 79 Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human
1664 genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

1665 80 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and
1666 genotype data to estimate haplotypes and unobserved genotypes. *Genetic*
1667 *epidemiology* **34**, 816-834, doi:10.1002/gepi.20533 (2010).

1668 81 Pei, Y. F., Zhang, L., Li, J. & Deng, H. W. Analyses and comparison of imputation-
1669 based association methods. *PloS one* **5**, e10827, doi:10.1371/journal.pone.0010827
1670 (2010).

1671 82 Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-
1672 analyses. *Nature protocols* **9**, 1192-1212, doi:10.1038/nprot.2014.071 (2014).

1673 83 Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-
1674 1004 (1999).

1675 84 Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genetic*
1676 *epidemiology* **35**, 809-822, doi:10.1002/gepi.20630 (2011).

1677 85 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from
1678 genome-wide association studies. *Bioinformatics (Oxford, England)* **32**, 1493-1501,
1679 doi:10.1093/bioinformatics/btw018 (2016).

1680 86 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links
1681 to Common Complex Disease. *Cell* **167**, 1415-1429.e1419,
1682 doi:10.1016/j.cell.2016.10.042 (2016).

1683 87 Benyamin, B. *et al.* Novel loci affecting iron homeostasis and their effects in
1684 individuals at risk for hemochromatosis. *Nature communications* **5**, 4926,
1685 doi:10.1038/ncomms5926 (2014).

1686 88 Binesh, N. & Rezghi, M. Fuzzy clustering in community detection based on
1687 nonnegative matrix factorization with two novel evaluation criteria. *Applied Soft*
1688 *Computing* **69**, 689-703 (2018).

1689 89 Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in
1690 Europeans. *Diabetes* **66**, 2888-2902, doi:10.2337/db16-1253 (2017).

1691 90 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell
1692 types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).

1693 91 Mikkelsen, T. S. *et al.* Comparative epigenomic analysis of murine and human
1694 adipogenesis. *Cell* **143**, 156-169, doi:10.1016/j.cell.2010.09.006 (2010).

1695 92 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
1696 characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

1697 93 Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on
1698 gene expression across human tissues. *Nature* **550**, 204-213,
1699 doi:10.1038/nature24277 (2017).

1700 94 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative
1701 trait loci in whole blood. *Nature genetics* **49**, 139-145, doi:10.1038/ng.3737 (2017).

1702 95 Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of
1703 known disease associations. *Nature genetics* **45**, 1238-1243, doi:10.1038/ng.2756
1704 (2013).

1705 96 Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait
1706 loci aids interpretation of genomic association studies. *Genome biology* **18**, 16,
1707 doi:10.1186/s13059-016-1142-6 (2017).

1708 97 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using
1709 predicted gene functions. *Nature communications* **6**, 5890,
1710 doi:10.1038/ncomms6890 (2015).

1711